**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ**

**РОССИЙСКОЙ ФЕДЕРАЦИИ**

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

(национальный исследовательский университет)»

**Институт (Филиал)** № 8 «Компьютерные науки и прикладная математика» **Кафедра**  806

**Группа** М8О-410Б-19  **Направление подготовки** 02.03.02 Фундаментальная информатика и информационные технологии

**Профиль**  Информатика и компьютерные науки

**Квалификация**  **бакалавр**

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

**БАКАЛАВРА**

На тему: Разработка рекомендательной системы на базе решения задачи стохастической оптимизации с использованием базы данных музея

Автор ВКРБ\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Шекунов Михаил Андреевич (\_\_\_\_\_\_\_\_\_\_)

(фамилия, имя, отчество полностью)

Руководитель Лемтюжникова Дарья Владимировна (\_\_\_\_\_\_\_\_\_\_)

(фамилия, имя, отчество полностью)

Консультант (\_\_\_\_\_\_\_\_\_\_)

(фамилия, имя, отчество полностью)

Консультант (\_\_\_\_\_\_\_\_\_\_)

(фамилия, имя, отчество полностью)

Рецензент (\_\_\_\_\_\_\_\_\_\_)

(фамилия, имя, отчество полностью)

**К защите допустить**

Заведующий кафедрой 806 Крылов Сергей Сергеевич (\_\_\_\_\_\_\_\_\_)

(№ каф) (фамилия, имя, отчество полностью)

\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2023г.

Москва 2023

**РЕФЕРАТ**

Выпускная квалификационная работа бакалавра состоит из 00 страниц, 00 рисунков, 00 таблиц, 00 использованных источников, 00 приложений.

Текст реферата должен отражать:

* + объект исследования или разработки;
  + цель работы;
  + методы или методологию проведения работы;­­­
  + результаты работы и их новизну;
  + область применения результатов;
  + рекомендации по внедрению или итоги внедрения результатов;
  + экономическую эффективность или значимость работы;
  + прогнозные предположения о развитии объекта исследования.

Если отчет не содержит сведений по какой-либо из перечисленных структурных частей реферата, то в тексте реферата она опускается, при этом последовательность изложения сохраняется.

Оптимальный объем текста реферата - 850 печатных знаков, но не более одной страницы машинописного текста.

В общее количество таблиц и рисунков входят и таблицы, и рисунки приложений.

СОДЕРЖАНИЕ

[ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ 4](#_Toc133903900)

[ВВЕДЕНИЕ 5](#_Toc133903901)

[1 ТЕОРИЯ 11](#_Toc133903902)

[1.1 Потребности в разработке гибридной рекомендательной системы на базе стохастической оптимизации для культурных учреждений 11](#_Toc133903903)

[1.2. Техническое оснащение и процесс сбора данных о посещении объектов в музее 12](#_Toc133903904)

[1.3 Анализ технических возможностей для реализации алгоритмов 13](#_Toc133903905)

[1.4 Анализ базы данных музея 14](#_Toc133903906)

[1.5 Поиск существующих рекомендательных систем 14](#_Toc133903907)

[1.6 Сопоставление имеющихся данных и существующих алгоритмов 15](#_Toc133903908)

[1.7 Коллаборативная фильтрация 17](#_Toc133903909)

[1.8 Система на основе схожести объектов 20](#_Toc133903910)

[1.9 Рекомендательная система, основанная на расстояниях между объектами 22](#_Toc133903911)

[1.10 Гибридная рекомендательная система 23](#_Toc133903912)

[1.11 Слияние результатов гибридной системы с результатами стохастической оптимизации 24](#_Toc133903913)

[ПРАКТИЧЕСКАЯ ЧАСТЬ 26](#_Toc133903914)

# **ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ**

В настоящей выпускной квалификационной работе бакалавра применяют следующие термины с соответствующими определениями:

Коллаборативная фильтрация — ­это один из методов построения прогнозов (рекомендаций) в рекомендательных системах, использующий известные предпочтения (оценки) группы пользователей для прогнозирования неизвестных предпочтений другого пользователя.

Стохастическая оптимизация — это метод оптимизации, используемый для решения задач, когда функционал, который нужно оптимизировать, не может быть вычислен точно или слишком сложен для аналитического решения.

# **ВВЕДЕНИЕ**

Актуальность темы данной работы связана с тем, что в настоящее время музеи играют важную роль в культурной жизни общества и привлекают множество посетителей, которые хотят ознакомиться с экспонатами, историей и культурой различных народов. Однако, посетители часто сталкиваются с проблемой ограниченного времени, что может вызвать у них чувство неудовлетворенности и привести к негативному опыту посещения музея.

Для решения данной проблемы, музеи начали применять технологии и методы, которые могут помочь посетителям более эффективно планировать свое время и выбирать интересующие экспонаты. В этом контексте, разработка рекомендательных систем становится все более актуальной и востребованной задачей, поскольку они могут помочь посетителям выбирать наиболее интересующие экспонаты и оптимизировать свое время.

Однако, существующие рекомендательные системы могут не учитывать все предпочтения посетителей и не предоставлять наиболее оптимальные рекомендации. В этой связи, разработка новых методов и алгоритмов рекомендательных систем с использованием современных методов стохастической оптимизации является важной задачей, которая может улучшить качество обслуживания посетителей и повысить удовлетворенность ими.

Разработка рекомендательной системы на базе решения задачи стохастической оптимизации для музея имеет высокую актуальность в контексте улучшения качества обслуживания посетителей и оптимизации использования времени на посещение экспонатов музея.

Таким образом, выполненная работа актуальна и с научно-методической/теоретической, и с практической точек зрения.

Цель данной работы - разработать рекомендательную систему на базе решения задачи стохастической оптимизации с использованием базы данных музея, которая будет предоставлять посетителям музея персонализированные рекомендации на основе их предпочтений и интересов. Рекомендательная система будет разработана на заказ винного музея во Франции и будет использоваться для помощи посетителям в выборе наиболее интересующих экспонатов, оптимизации использования времени на посещение музея и повышения удовлетворенности посетителей.

Для достижения поставленной цели в работе были решены следующие задачи:

1. изучение существующих методов и алгоритмов рекомендательных систем и выбор оптимальных для решения задачи,
2. преобразование базы данных под рекомендательную систему,
3. реализация нескольких типов рекомендательных систем,
4. разработка объединения результатов работы нескольких алгоритмов рекомендательных систем,
5. разработка компонентов рекомендательной системы, с помощью которых учитываются результаты решения задачи стохастического программирования по определению маршрута движения туриста с учетом различных сценариев,
6. реализация и тестирование рекомендательной системы, результатом которой является близкий к оптимальному маршрут движения туриста.

Для выполнения работы были использованы различные инструменты и методы, включая:

1. Язык программирования Python. Python является одним из наиболее распространенных языков программирования, который широко используется для научных вычислений, машинного обучения и анализа данных. В данной работе Python использовался для разработки алгоритмов и реализации рекомендательной системы.
2. Файлы CSV: Формат CSV (Comma Separated Values) является одним из наиболее распространенных форматов хранения и обмена данными. В данной работе файлы CSV использовались для хранения и обработки данных, связанных с экспонатами, посетителями музея, результатов работы рекомендательных систем.
3. Библиотека SciPy: SciPy — это библиотека для научных вычислений в Python, которая содержит множество инструментов для решения задач оптимизации, линейной алгебры, статистического анализа и другого. В данной работе библиотека SciPy использовалась для реализации алгоритмов стохастической оптимизации.
4. Иные библиотеки Python для работы с данными. В данной работе использовались различные библиотеки Python для работы с данными, такие как библиотека os для работы с файловой системой, библиотека Pandas для анализа данных, библиотека NumPy для научных вычислений и так далее.

В рамках данной работы были получены следующие основные результаты:

1. Отсортированные рекомендации с баллами для трех алгоритмов рекомендательных систем:
2. Коллаборативная фильтрация: Для каждого посетителя музея на основе его предыдущих посещений формируется список рекомендаций, который содержит объекты, наиболее похожие на те, которые посетитель уже посмотрел. После этого список рекомендаций сортируется по баллам, которые рассчитываются на основе сходства между объектами и предпочтений посетителя.
3. Рекомендательная система, основанная на схожести объектов: Для каждого посетителя музея на основе его предыдущих посещений формируется список рекомендаций, который содержит объекты, наиболее похожие на те, которые посетитель уже посмотрел. После этого список рекомендаций сортируется по баллам, которые рассчитываются на основе сходства между объектами.
4. Рекомендательная система, основанная на расстояниях: Для каждого посетителя музея на основе его предыдущих посещений формируется список рекомендаций, который содержит объекты, находящиеся ближе всего к тем, которые посетитель уже посмотрел. После этого список рекомендаций сортируется по баллам, которые рассчитываются на основе расстояний между объектами.
5. Отсортированный результат слияния рекомендательных систем по баллам. Для каждого посетителя музея получены списки рекомендаций на основе трех различных алгоритмов рекомендательных систем. После этого списки рекомендаций объединены и отсортированы по баллам, полученным на основе результатов трех алгоритмов.
6. Итоговый отсортированный список рекомендаций: Для каждого посетителя музея формируется индивидуальный список рекомендаций, который учитывает результат слияния трех алгоритмов рекомендательных систем и работу эвристического алгоритма стохастической оптимизации. Список рекомендаций отсортирован по баллам, которые рассчитываются на основе сходства между объектами и предпочтений посетителя, а также результатов работы стохастической оптимизации. Кроме того, данные для каждого посетителя формируются индивидуально после посещения нового объекта, что позволяет рекомендовать экспонаты, наиболее соответствующие интересам и предпочтениям каждого конкретного посетителя.

Все результаты были записаны в файлы CSV, которые могут быть использованы для последующего анализа и улучшения алгоритмов рекомендательной системы.

Результаты работы предназначены для внедрения в работу музея с целью улучшения качества обслуживания посетителей и оптимизации времени, затрачиваемого на просмотр экспонатов. Разработанный алгоритм рекомендации может быть использован музеем для предоставления индивидуальных рекомендаций посетителям, исходя из их предпочтений и интересов. Полученные результаты могут быть использованы для улучшения качества обслуживания и увеличения удовлетворенности посетителей, что, в свою очередь, может привести к увеличению посещаемости музея.

Кроме того, результаты работы могут быть использованы для дальнейших исследований в области рекомендательных систем и оптимизации процессов в музеях. Полученные результаты могут послужить основой для разработки более эффективных алгоритмов рекомендации и оптимизации процессов в музеях, что в свою очередь может привести к повышению качества обслуживания и удовлетворенности посетителей.

Также результаты работы могут быть использованы для интеграции с другими системами управления музеем, такими как системы бронирования билетов, системы управления посетителями и системы анализа посещаемости. Это может помочь музею снизить затраты на управление и повысить эффективность работы, что в итоге может привести к улучшению качества обслуживания посетителей и удовлетворенности их потребностей.

Использование разработанного алгоритма рекомендации на базе решения задачи стохастической оптимизации с использованием базы данных музея позволяет музею сократить время, затрачиваемое посетителями на поиск и выбор интересных для них экспонатов. Благодаря предоставлению индивидуальных рекомендаций на основе предпочтений и интересов посетителей, музей может увеличить удовлетворенность посетителей и улучшить качество обслуживания. Посетители будут получать персонализированные рекомендации, которые помогут им сэкономить время и сделать посещение музея более приятным и интересным.

Кроме того, использование разработанной рекомендательной системы может привести к повышению посещаемости музея, так как посетители будут получать рекомендации о тех экспонатах, которые наиболее соответствуют их интересам, что может заинтересовать их посетить музей снова. Это может привести к увеличению прибыли музея и повышению его популярности среди посетителей.

Также использование разработанной рекомендательной системы может помочь музею сократить время на обслуживание посетителей и уменьшить нагрузку на персонал музея. Благодаря индивидуальным рекомендациям, посетители смогут быстрее найти интересные для них экспонаты и не будут требовать помощи персонала музея в выборе экспонатов.

В целом, использование разработанной рекомендательной системы позволяет музею улучшить качество обслуживания посетителей, повысить эффективность работы и повысить прибыль.

# **1 ТЕОРИЯ**

## 1.1 Потребности в разработке гибридной рекомендательной системы на базе стохастической оптимизации для культурных учреждений

В современном мире наблюдается быстрый рост информации, что создает огромные проблемы для пользователей в выборе наиболее подходящей информации. Одним из решений этой проблемы являются рекомендательные системы, которые предоставляют пользователям персонализированные рекомендации на основе их предпочтений и интересов. Рекомендательные системы широко используются в различных областях, включая электронную коммерцию, социальные сети, видео- и музыкальные сервисы и т.д. Однако, разработка рекомендательных систем, которые бы удовлетворяли требованиям пользователей, остается актуальной и сложной задачей.

В сфере культуры и искусства рекомендательные системы также играют важную роль. Музеи, театры и другие культурные учреждения сталкиваются с проблемой обеспечения качественного обслуживания посетителей и увеличения их удовлетворенности. Рекомендательные системы могут помочь решить эту проблему, предоставляя посетителям персонализированные рекомендации посещения экспонатов или мероприятий на основе их интересов и предпочтений.

Однако, существующие рекомендательные системы не всегда удовлетворяют потребности пользователей. Например, некоторые системы не учитывают индивидуальные потребности и предпочтения пользователей, а другие могут оказаться неэффективными в силу недостаточной точности предсказания рекомендаций.

Именно поэтому разработка новых методов и алгоритмов для создания эффективных рекомендательных систем остается актуальной и важной задачей. В рамках данной работы рассматривается разработка гибридной рекомендательной системы на базе решения задачи стохастической оптимизации с использованием базы данных музея. Она позволит музею предоставлять посетителям персонализированные рекомендации на основе их интересов и предпочтений, что, в свою очередь, улучшит качество обслуживания и повысит удовлетворенность.

## 1.2. Техническое оснащение и процесс сбора данных о посещении объектов в музее

В музее, из которого был заказ, используются планшетники для предоставления посетителям дополнительной информации об экспонатах. Для получения информации посетитель должен отсканировать QR-код, который находится на каждом экспонате, с помощью камеры планшетника. После сканирования QR-кода на экране планшета появляется информация об объекте, такая как его название, дата создания, описание, автор и т.д. Процесс сканирования QR-кода на планшете для посетителей музея довольно простой. Для того, чтобы получить дополнительную информацию об экспонате, посетитель должен с помощью камеры планшета отсканировать QR-код, который находится на каждом объекте. После этого на экране планшета появляется информация об объекте, которую посетитель может изучать.

Однако, для разработчиков системы, использование QR-кодов также предоставляет возможность собирать данные о поведении посетителей. Система может отслеживать, какие объекты были посещены каждым посетителем, как долго они останавливались на каждом объекте и когда они переходили к следующему. Эти данные могут быть использованы для анализа интересов посетителей, и для того, чтобы оптимизировать предложения рекомендаций на основе этого поведения.

Кроме того, система также может собирать данные о посещенных объектах, которые могут быть использованы музеем для анализа посещаемости и популярности различных экспонатов. Эта информация может быть полезна для планирования выставок, реконструкции залов и других мероприятий, связанных с улучшением посещаемости музея.

Благодаря простоте процесса сканирования QR-кодов на планшетах в музее, нам не нужно просить пользователей оставлять оценки или комментарии об объектах, что упрощает взаимодействие с системой и делает процесс более естественным. Тем не менее, во время сканирования мы можем записывать время посещения каждого объекта, и использовать эту информацию в качестве оценки интереса посетителя к конкретному экспонату. Это может быть полезно при разработке рекомендательной системы, которая будет учитывать предпочтения и интересы каждого посетителя, а также оптимизировать их посещение музея. В свою очередь, такая система может помочь музею повысить качество обслуживания, улучшить удовлетворенность посетителей и повысить посещаемость.

Таким образом, использование QR-кодов в музее позволяет не только предоставлять посетителям дополнительную информацию об экспонатах, но и собирать данные о поведении посетителей, которые могут быть использованы для улучшения качества обслуживания и оптимизации рекомендаций.

## 1.3 Анализ технических возможностей для реализации алгоритмов

Для работы алгоритмов мы используем слабопроизводительные устройства, так как это планшеты, которые обычно имеют ограниченные ресурсы. В связи с этим, мы должны снизить количество вычислений, которые производятся на устройствах, и использовать максимально доступные данные для получения оценок пользователя. В данном случае мы используем время, проведенное пользователем у каждого объекта в качестве такой оценки.

Следует отметить, что музей является большим и может быть ограничен доступ к интернету или локальной сети. Это означает, что все вычисления для рекомендательной системы должны выполняться на планшетах. Однако мы должны обеспечивать при этом высокую точность рекомендаций.

## 1.4 Анализ базы данных музея

База данных музея содержит информацию о посетителях, которые используют планшеты для сканирования QR-кодов экспонатов и получения информации о них. Из этих данных мы извлекли информацию о времени посещения каждого экспоната каждым посетителем. Также в базе данных содержится информация об экспонатах, включая их название, категорию, описание и номер в базе данных. Мы использовали эту информацию для создания csv файлов с отобранными посетителями и экспонатами.

Кроме того, в базе данных содержится информация о расположении экспонатов в помещении, включая номер зала и номер экспоната, а также информация о категориях помещений, к которым принадлежат экспонаты.

## 1.5 Поиск существующих рекомендательных систем

Алгоритмы рекомендательных систем могут быть различными и, как правило, определяются конкретной задачей, которую необходимо решить. В литературе (<https://doi.org/10.1023/A:1021240730564>) существует множество алгоритмов, которые можно использовать для построения рекомендательных систем. Некоторые из них включают:

1. Content-based filtering (CBF) - основан на анализе характеристик продуктов или услуг, которые пользователь оценил или потребовал.
2. Collaborative filtering (CF) - основан на схожести пользователей, исходя из их предпочтений. Может быть двух типов: user-based и item-based.
3. Knowledge-based filtering - использует экспертную систему для генерации рекомендаций.
4. Hybrid recommender systems - комбинируют различные алгоритмы, такие как CBF и CF, чтобы улучшить качество рекомендаций.
5. Demographic-based filtering - использует информацию о демографических характеристиках пользователей, таких как возраст, пол, место жительства и т.д.
6. Knowledge-based filtering with demographic information - комбинирует информацию о демографии пользователей с экспертной системой для генерации рекомендаций.
7. Context-aware filtering - учитывает контекст, в котором пользователь просматривает или потребляет продукты, например, местоположение, время и т.д.
8. Social-based filtering - основан на анализе социальных связей пользователей и используется для генерации рекомендаций, основанных на социальных связях.

## 1.6 Сопоставление имеющихся данных и существующих алгоритмов

При разработке рекомендательных систем, одним из важных аспектов является учет демографических характеристик пользователей. Эти данные могут включать в себя возраст, пол, местоположение и т.д. Однако, сбор такой информации может стать проблемой, особенно в контексте музеев, где посетители могут заходить на короткий срок и не захотеть тратить время на заполнение анкет или опросников. Дополнительно, вопросы о демографических характеристиках могут вызвать недовольство и неудовольствие у пользователей, особенно если они считают, что эти вопросы не имеют никакого отношения к их посещению музея. Из-за этих факторов, сбор демографических данных может стать непрактичным и неэффективным для использования в рекомендательных системах.

В связи с этим, мы решили использовать две системы, которые могут работать без дополнительных данных от пользователя: коллаборативную фильтрацию и рекомендательную систему, основанную на сходстве объектов.

Коллаборативная фильтрация основана на анализе предпочтений пользователей и сопоставлении их с предпочтениями других пользователей, чтобы предложить рекомендации.

Рекомендательная система, использующая сходство, основана на анализе характеристик объектов и сопоставлении их с предпочтениями пользователей.

Обе системы не требуют от пользователя дополнительных данных, поэтому возможно использование имеющиеся данные из базы данных музея.

Для использования коллаборативной фильтрации нужно определить, какие пользователи могут быть схожими по предпочтениям и на основе этого определить, какие экспонаты можно порекомендовать. Для основанной на сходстве рекомендательной системы нам нужно определить характеристики объектов, которые будут использоваться для сопоставления с предпочтениями пользователей.

База данных музея содержит информацию о посетителях и экспонатах, которые могут быть использованы для выбора рекомендательных систем. Были собраны данные о времени, проведенном посетителями у каждого экспоната, и использовали эту информацию для определения, какие экспонаты могут быть наиболее интересны для посетителей. Также была получена информация о характеристиках объектов и их расположении в помещениях, что позволяет определить, какие экспонаты могут быть связаны друг с другом и могут быть интересны для пользователей, основываясь на их времени пребывания в музее.

Поскольку мы имеем информацию о принадлежности экспонатов к определенным комнатам и их расположении внутри них, возможно использование этой информации для формирования рекомендаций, которые будут основываться на расстоянии между объектами. Поэтому дополнительно включаем в разработку рекомендательной системы алгоритм, который учитывает расстояния между объектами в музее. Таким образом, пользователи получат рекомендации по экспонатам, находящимся в близости друг от друга, что может быть особенно удобно для тех, кто хочет изучить определенную тему или эпоху.

Для учета расстояний между объектами можно использовать данные о фактическом расположении экспонатов в музее, которые были извлечены из базы данных музея. Таким образом, можно более точно определить, какие объекты находятся ближе друг к другу, и использовать эту информацию для формирования рекомендаций.

## 1.7 Коллаборативная фильтрация

Коллаборативная фильтрация - одна из наиболее распространенных рекомендательных систем, используемых для предсказания рейтингов или предпочтений пользователей. Работа этой системы основана на предположении, что если два пользователя оценили некоторый набор объектов (фильмов, музыки, товаров и т.д.) схожим образом, то они будут иметь схожие вкусы и интересы.

В подходе, анализирующем пользователей, сначала вычисляется сходство между всеми парами пользователей, используя меры сходства, такие как косинусное сходство или корреляция Пирсона. Затем, для каждого пользователя находятся похожие на него пользователи, и вычисляются взвешенные оценки объектов на основе их оценок и сходства с выбранными объектами. Эти оценки затем используются для формирования рекомендаций.

Коллаборативная фильтрация имеет ряд преимуществ, таких как возможность учитывать неявные предпочтения пользователей.

Для реализации коллаборативной системы необходимо выполнить несколько шагов. Сначала нужно получить данные о предпочтениях пользователей, то есть оценки, которые они выставляют объектам. Затем эти данные нужно преобразовать в матрицу оценок. Обычно матрица оценок имеет вид таблицы, где каждая строка соответствует пользователю, каждый столбец – объекту, а ячейка таблицы содержит оценку, которую пользователь выставил объекту. Схематично данная рекомендательная система и матрица оценок представлены на рисунке 1.

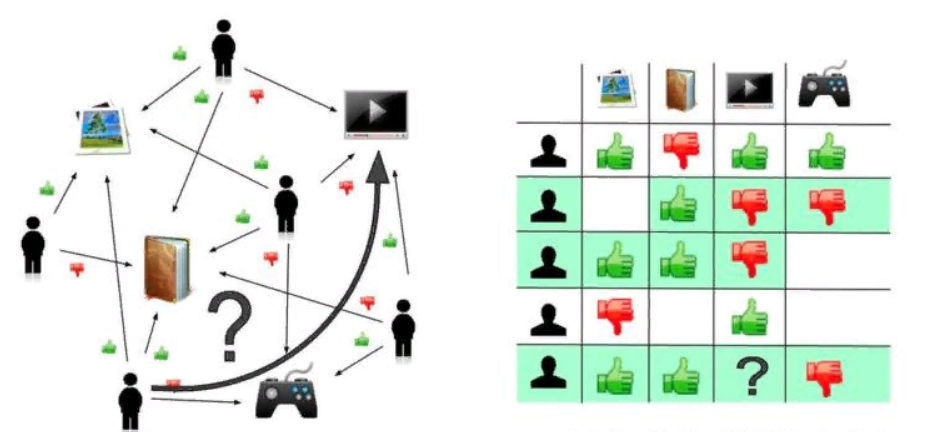


Рисунок 1 – Иллюстрация коллаборативной фильтрации

После этого можно перейти к этапу поиска похожих пользователей или объектов. Для этого используются различные алгоритмы, такие как метод k-ближайших соседей, сингулярное разложение и другие. Они позволяют определить наиболее похожие пользователей или объекты на основе сравнения их оценок.

Далее необходимо выполнить этап предсказания оценок. Для этого используются различные методы, такие как среднее значение оценок соседей или взвешенное среднее, где веса зависят от степени похожести соседей. На основе предсказанных оценок можно формировать список рекомендаций.

Важным моментом является оценка качества рекомендаций. Для этого используются различные метрики, такие как точность, полнота, F-мера и др. Также проводятся эксперименты на тестовых данных, чтобы оценить эффективность системы.

Одним из недостатков коллаборативной системы является холодный старт, то есть невозможность предложить рекомендации для новых пользователей или объектов. Также система может страдать от проблемы рекомендации известных объектов, то есть предлагать уже известные пользователю объекты.

Коллаборативная фильтрация эффективна, потому что она основана на анализе предпочтений пользователей и нахождении сходств между их предпочтениями. Это позволяет системе предлагать пользователю объекты, которые ему могут понравиться, на основе предпочтений их похожих пользователей.

Исследования показывают, что коллаборативная фильтрация является одной из самых эффективных и точных методов рекомендаций. Например, в статье "A Survey of Collaborative Filtering Techniques" (1997) <https://doi.org/10.1155/2009/421425> был проведен обзор и сравнение различных методов коллаборативной фильтрации, и результаты показали, что они превосходят другие методы рекомендаций по точности.

Также, исследование "Collaborative Filtering Recommender Systems" (2011) [Collaborative Filtering Recommender Systems | Now Foundations and Trends books | IEEE Xplore](https://ieeexplore.ieee.org/document/8187128/) выявило, что коллаборативная фильтрация имеет высокую точность и показывает хорошие результаты в рекомендациях в различных областях, включая музыку, фильмы, книги и т.д.

Из этих исследований следует, что коллаборативная фильтрация является эффективным методом рекомендаций, особенно в случае большого количества пользователей и объектов, что делает ее подходящим выбором для нашей задачи рекомендации экспонатов в музее.

## 1.8 Система на основе схожести объектов

Система рекомендаций на основе сходства объектов (item-based recommendation system) предполагает, что если пользователь понравился один объект, то ему могут понравиться и другие объекты, которые похожи на этот объект. Для построения такой системы сначала нужно создать матрицу сходства между объектами. Это можно сделать с помощью различных метрик сходства, таких как косинусное сходство или евклидово расстояние.

После того, как матрица сходства создана, для каждого объекта можно найти его похожие объекты. Например, если пользователь заинтересован в экспонате, который относится к разделу «История технологий», то ему могут быть интересны и другие экспонаты из этого же раздела. Для этого для каждого объекта находятся наиболее похожие на него объекты из того же раздела, после чего список объектов сортируется по убыванию степени сходства. Пользователю предлагаются наиболее похожие объекты из этого списка.

Основное преимущество системы рекомендаций на основе сходства объектов заключается в том, что она может предложить пользователю объекты, которые ему могут понравиться, но которые он сам бы не нашел. При этом не требуется большого количества информации о пользователе, достаточно знать, какой объект ему понравился, чтобы на основе сходства объектов подобрать другие, которые могут его заинтересовать.

Однако система рекомендаций на основе сходства объектов имеет и свои недостатки. В частности, она не учитывает индивидуальные предпочтения пользователя и его контекст, такой как время, место и настроение. Также может возникнуть проблема "холодного старта", когда новый объект, который еще не имеет достаточного количества оценок и связей с другими объектами, не будет учтен системой.

Для решения этих проблем, в систему можно добавить дополнительные факторы, такие как временные, географические или социальные данные. Это позволяет учитывать контекст и индивидуальные предпочтения пользователя, а также повышает качество рекомендаций.

Исследования показывают, что системы, основанные на сходстве объектов, могут быть эффективными в рекомендации объектов, которые могут заинтересовать пользователя. Например, исследование "Content-Based Recommender Systems: State of the Art and Trends" от 2014 года показало, что системы, основанные на сходстве объектов, могут быть эффективными в рекомендации музейных экспонатов, книг и фильмов.

Другое исследование "A Survey of Content-based Recommender Systems Based on Machine Learning Techniques" от 2016 года также подтвердило эффективность систем, основанных на сходстве объектов, и показало, что они могут быть эффективными в рекомендации объектов в различных областях, включая музеи и галереи.

Таким образом, системы, основанные на сходстве объектов, могут быть эффективными в рекомендации объектов, которые могут заинтересовать пользователя. Данная рекомендательная система проиллюстрирована на рисунке номер 2.



Рисунок 2 – Content-based рекомендательная система

## 1.9 Рекомендательная система, основанная на расстояниях между объектами

Рекомендательная система, основанная на расстояниях, использует информацию о том, как далеко находятся объекты друг от друга в пространстве музея, чтобы рекомендовать пользователю близлежащие экспонаты. Данная система позволяет более точно учитывать пространственную структуру музея, что может быть полезным в случаях, когда пользователь желает ознакомиться с объектами в конкретном порядке, или имеется маршрут, оптимизированный для посещения объектов в определенном порядке.

Основным алгоритмом данной системы является алгоритм k-ближайших соседей (k-nearest neighbors, KNN). Он основан на том, что объекты, расположенные близко друг к другу в пространстве музея, вероятно, будут интересны пользователю, если пользователь уже проявил интерес к некоторым объектам. В этом случае алгоритм выбирает k ближайших экспонатов к тем, которые уже были посещены пользователем, и предлагает их как рекомендации.

## 1.10 Гибридная рекомендательная система

Каждая из рекомендательных систем имеет свои сильные и слабые стороны. Коллаборативная фильтрация работает на основе схожести предпочтений пользователей, а рекомендация на основе сходства объектов использует сходство между объектами для выдачи рекомендаций. Рекомендательная система на основе расстояний между объектами использует расстояние между объектами для выдачи рекомендаций.

Одним из преимуществ объединения результатов рекомендательных систем является то, что это позволит снизить риск ошибок. Каждая из систем может допустить ошибку, что может привести к неверным рекомендациям. Однако объединение результатов позволит увеличить точность рекомендаций и снизить риск ошибок. Также это помогает нивелировать ошибки, вызванные недостаточным количеством данных или искажением результатов в одной из систем.

Кроме того, объединение результатов может увеличить разнообразие рекомендаций. Каждая из систем может выдавать рекомендации, которые схожи между собой. Однако, объединение результатов позволит увеличить разнообразие рекомендаций и предложить пользователю более широкий выбор.

Одним из главных плюсов такого подхода является уменьшение проблемы "холодного старта", то есть ситуации, когда рекомендательная система не может предложить рекомендации для нового посетителя, у которого еще нет истории посещений.

Объединение результатов также позволит учесть различные критерии рекомендаций, которые могут быть важны для разных пользователей. Например, один пользователь может предпочитать получать рекомендации на основе сходства объектов, а другой – на основе схожести предпочтений. Объединение результатов позволит учесть все эти критерии и выдать наиболее подходящие рекомендации для каждого пользователя.

Такая система может обеспечивать более точную рекомендацию для активных пользователей, так как учитывает их предпочтения, а также оценки, которые они давали объектам в прошлом.

В результате объединения в гибридную всех трех систем - коллаборативной фильтрации, системы на основе сходства объектов и системы на основе расстояний, мы получаем комплексную рекомендательную систему, которая учитывает разные аспекты предпочтений посетителей и повышает точность рекомендаций.

## 1.11 Слияние результатов гибридной системы с результатами стохастической оптимизации

Слияние результатов гибридной рекомендательной системы с эвристикой стохастической оптимизации является еще одним шагом в улучшении качества рекомендаций. Главным преимуществом использования стохастической оптимизации в качестве эвристики является то, что она может помочь в поиске оптимального решения, основываясь на статистических методах и случайных величинах.

Одним из главных вызовов при слиянии результатов гибридной системы с эвристикой является то, что стохастическая оптимизация не дает точных баллов для каждого объекта. Она работает с вероятностями и распределениями, что может быть не очень удобным для смешивания с другими рекомендательными системами, которые работают с точными оценками.

Чтобы решить эту проблему, мы можем использовать различные методы смешивания результатов, чтобы дать больший вес объектам, которые были лучше оценены в стохастической оптимизации. Мы можем также установить пороговое значение, при котором объекты с низкими оценками из стохастической оптимизации будут исключены из списка рекомендаций.

Важным моментом при слиянии результатов является также учет взаимодействия между пользователями и объектами. В случае стохастической оптимизации, каждый пользователь может получить свой собственный набор рекомендаций, основанный на его предпочтениях и интересах. При слиянии результатов, мы должны учитывать эти индивидуальные различия и попытаться дать каждому пользователю наиболее подходящие рекомендации.

# **2 ПРАКТИЧЕСКАЯ ЧАСТЬ**

**2.1 Используемый язык программирования**

Для создания рекомендательной системы был выбрал язык программирования Python, так как он является одним из наиболее подходящих языков программирования для решения данной задачи. Выбор языка программирования Python для данной работы обусловлен рядом его преимуществ, которые делают его наиболее подходящим инструментом для разработки рекомендательной системы на базе стохастической оптимизации.

Простота и интуитивность синтаксиса Python делает его доступным для использования в задачах программирования разного уровня сложности, что особенно важно в контексте данной работы.

Кросплатформенность Python позволяет использовать его на любой операционной системе, что облегчает работу с данными на различных устройствах. Так рекомендательная система не зависит от операционной системы планшета, на которой она будет запущена.

Наличие обширной коллекции библиотек и модулей Python, таких как NumPy, Pandas, Scikit-learn и os, делает его идеальным инструментом для работы с данными, машинного обучения и создания рекомендательных систем.

Поддержка функционального программирования Python делает код более простым и структурированным, что особенно важно для работы над сложными проектами.

Высокая скорость выполнения кода благодаря компиляции некоторых частей Python делает его подходящим выбором для задач, связанных с вычислительными процессами и обработкой больших объемов данных.

Интерактивная консоль Python позволяет быстро тестировать и отлаживать код, что ускоряет процесс разработки.

Простая и эффективная система управления пакетами Python, такая как pip, позволяет удобно устанавливать и управлять сторонними библиотеками, что также сильно упрощает разработку рекомендательной системы на базе стохастической оптимизации.

## 2.2 Используемые библиотеки и модули

Для успешной реализации рекомендательной системы были использованы различные библиотеки и инструменты. Краткое описание и обоснование использование библиотек приведено ниже.

Библиотека SciPy предоставляет множество функций и инструментов для научных и инженерных вычислений, включая построение матриц сходств. Одно из основных преимуществ SciPy - это скорость работы. Она оптимизирована для вычислительных задач и предоставляет высокую производительность благодаря использованию языка программирования C. Более того, SciPy имеет богатую коллекцию функций для работы с различными типами данных и предоставляет мощный и гибкий функционал для настройки и управления вычислениями.

Библиотека os является стандартной библиотекой Python, которая предоставляет множество функций для работы с операционной системой, такие как чтение и запись файлов, управление директориями и многое другое. Одно из основных преимуществ os - это его простота использования. Функции этой библиотеки имеют понятный синтаксис и предоставляют широкий спектр возможностей для работы с файловой системой. Кроме того, библиотека os позволяет писать переносимый код, который будет работать на разных операционных системах.

Библиотека pandas является одной из самых популярных библиотек для работы с данными в Python. Ее основное преимущество - это возможность работы с файлами csv и другими форматами данных. Pandas предоставляет мощный и гибкий функционал для чтения и записи данных, фильтрации, агрегации и многих других операций над ними. Одним из ключевых преимуществ pandas является скорость работы. Она оптимизирована для быстрой обработки больших объемов данных и предоставляет множество инструментов для ускорения вычислений. Кроме того, pandas предоставляет удобный и интуитивно понятный синтаксис, который делает ее доступной для использования даже новичкам в области анализа данных.

Библиотека numpy является одной из самых популярных библиотек для работы с массивами и матрицами в Python. Ее основное преимущество - это скорость работы и мощный функционал для работы с многомерными массивами данных. Numpy использует оптимизированный C-код для выполнения матричных операций, что позволяет ей работать очень быстро и обрабатывать большие объемы данных. Кроме того, numpy предоставляет множество инструментов для работы с матрицами, векторами, линейной алгеброй, численными методами и многими другими математическими операциями. Одним из дополнительных преимуществ numpy является ее способность работать с различными типами данных, включая числа с плавающей точкой, комплексные числа и другие. Кроме того, numpy имеет мощный и удобный интерфейс для интеграции с другими библиотеками, такими как pandas и matplotlib, что делает ее очень гибким и мощным инструментом для работы с данными и анализа данных.