

МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической теории  
упругости и биомеханики

**Проектирование рекомендательной системы на основе близости  
предпочтений**

**АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студента 4 курса 442 группы

направления 09.03.03 – Прикладная информатика

механико-математического факультета

Мраморнова Андрея Константиновича

Научный руководитель

к. ю. н., доцент

\_\_\_\_\_

подпись, дата

Р. В. Амелин

Зав. кафедрой

д. ф.-м. н., профессор

\_\_\_\_\_

подпись, дата

Л.Ю. Коссович

## **Введение**

Рекомендательные системы - это интеллектуальные системы, которые предлагают пользователю контент исходя из его предпочтений. В последнее время все большее количество компаний начинают использовать данные алгоритмы подбора информации. Среди компаний, флагманов рынка, которые могут предложить наилучший алгоритм интеллектуальной системы, мы можем назвать таких гигантов как Netflix, Spotify, Apple, Google и даже Amazon.

Учитывая большое количество компаний, которые включают в свои продукты рекомендательные системы, можно с уверенностью сказать, что данная тема является очень актуальной на сегодняшний день.

Работа состоит из введения, трех разделов и заключения. Объем работы – 44 страницы и 39 рисунков.

Цель данной работы - разбор всех видов рекомендательных систем, их плюсы, минусы и проблемы, с которыми они могут столкнуться. Но главной задачей является проектирование рекомендательной системы на основе близости предпочтений. В качестве объекта проекта есть фильмы, данные о которых мы используем и анализируем. Суть проекта заключается в том, что система, выдает рекомендации по определенному фильму, на основе оценок пользователей, что оценивали и этот фильм тоже.

Результат проекта - список рекомендованных фильмов на основе близости предпочтений.

## **Основная часть**

Как и многие методы машинного обучения, рекомендательная система делает прогнозы на основе исторического поведения пользователей. В частности, он предназначен для прогнозирования предпочтений пользователя и для набора элементов на основе прошлого опыта.

Рекомендательные системы сейчас настолько распространены, что многие из нас используют их, даже не подозревая об этом. Поскольку мы не можем просматривать все продукты или контент на веб-сайте, система рекомендаций играет важную роль, помогая нам улучшить взаимодействие с пользователем, а

также предоставляя нам больше ресурсов, благодаря которым мы можем экономить свои ресурсы.

Некоторые примеры действующих систем рекомендаций включают рекомендации по продуктам на Amazon, рекомендуемые видео на YouTube, музыку на Spotify, новостную ленту Facebook и Google Ads.

Компонентом любой из этих систем является рекомендательная функция, которая берет информацию о пользователе и прогнозирует, например, рейтинг, который пользователь может присвоить продукту. Прогнозирование оценок пользователей, даже до того, как пользователь их фактически предоставил, делает рекомендательные системы мощным инструментом.

### **Принципы работы рекомендательных систем**

Первое, что нужно рассмотреть – это понимание взаимосвязей. Взаимосвязи или отношения – предоставляют рекомендательным системам понимание клиентов. Есть три основных типа:

- Взаимосвязь между пользователем и продуктом

Возникают, когда некоторые пользователи имеют привязанность или предпочтение к конкретным продуктам, которые им нужны. Например, игрок в баскетбол может отдавать предпочтение предметам, связанным с баскетболом, поэтому веб-сайт будет строить отношения между пользователем и продуктом по типу «игрок – баскетбол».

- Взаимосвязь продукт-продукт

Взаимосвязь продукт-продукт возникают, когда элементы похожи по своей природе либо по внешнему виду, либо по описанию. Некоторые примеры включают книги или музыку того же жанра, блюда той же кухни или новостные статьи с определенного события.

- Взаимосвязь пользователь-пользователь

Взаимосвязь между пользователем и пользователем возникают, когда некоторые клиенты имеют схожие вкусы по отношению к конкретному продукту

или услуге. Примеры включают общих друзей, схожее происхождение, схожий возраст и т. д.

Помимо взаимосвязей, рекомендательные системы используют следующие типы данных:

- Данные о поведении пользователей

Данные о поведении пользователей – это полезная информация о взаимодействии пользователя с продуктом. Его можно собрать из рейтингов, кликов и истории покупок.

- Демографические данные пользователей

Демографическая информация пользователя связана с личной его информацией, такой как возраст, образование, доход и местонахождение.

- Данные атрибутов продукта

Атрибуты продукта – это информация, относящаяся к самому продукту, например, жанр в случае книг, актерский состав в случае фильмов, кухня в случае еды

Данные могут быть предоставлены разными способами. Есть два особенно важных метода: явная и неявная оценка.

- Явные рейтинги

Явные оценки, которые предоставляются пользователем. Они делают вывод о предпочтениях пользователя. Примеры включают рейтинги, обзоры, отзывы, лайки и подписку. Поскольку пользователи не всегда оценивают продукты, получить точные оценки может быть сложно.

- Скрытые рейтинги

Неявный рейтинг предполагает косвенные предпочтения пользователей. Они позволяют сделать вывод о поведении пользователя, и их легко получить, поскольку пользователи подсознательно щелкают мышью. Примеры включают клики, просмотры, прослушивания и покупки.

Рассмотрим пару типов сходства, на основе которых выполняется фильтрация по товарам или пользователям.

- Сходство продукта (фильтрация товаров)
- Сходство пользователей (фильтрация пользователей и пользователей)

### Типы рекомендательных систем

Рекомендательные системы подразделяются на некоторые типы, мы изобразили их в виде схемы в соответствии с рисунком 1.

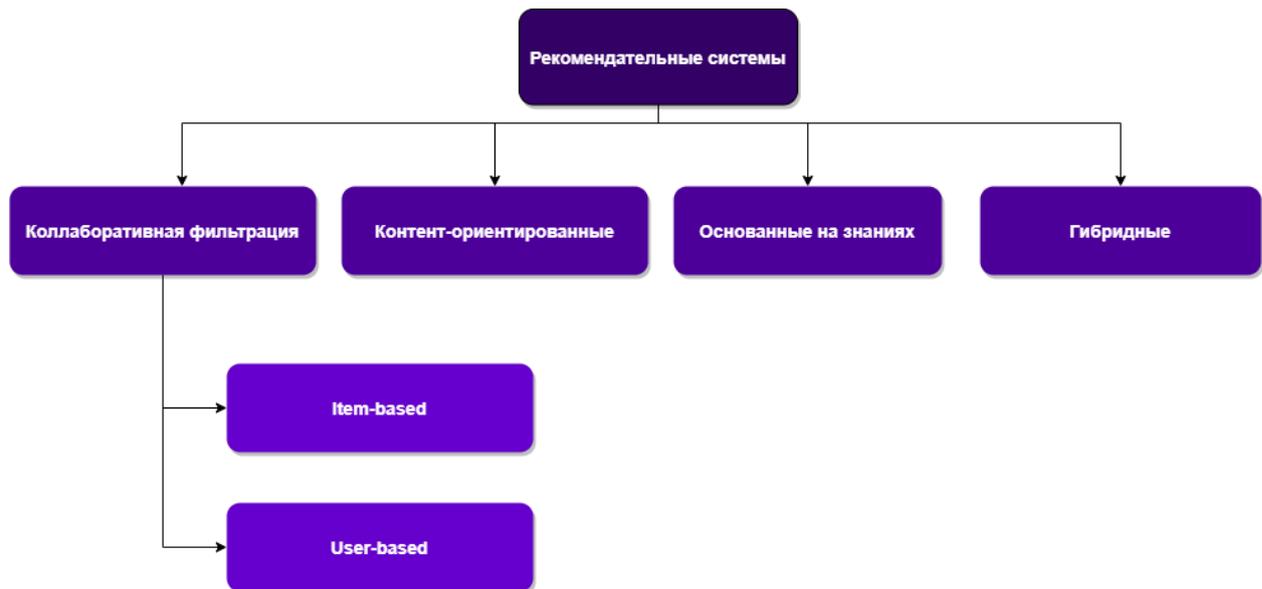


Рисунок 1 – Типы рекомендательных систем

Далее мы подробнее поговорим о каждом из типов рекомендательных систем.

- Контент-ориентированные

Этот тип рекомендательной системы использует дополнительные сведения об элементах и пользователях для создания предложений по элементам, атрибутам. Например, система рекомендаций для фильма может анализировать дополнительную информацию, такую как жанры фильмов, основные действующие лица и продолжительность фильма.

Системы рекомендаций, основанные на содержании, используют свои знания о каждом продукте, чтобы рекомендовать новые. Рекомендательные системы, основанные на содержании, работают хорошо, когда данные о содержании предоставляются заранее. “Сходство” измеряется по атрибутам продукта.

- Коллаборативная фильтрация

Коллаборативная или же совместная фильтрация. Этот тип рекомендательной системы использует рейтинг пользователей для выработки рекомендаций. Он также использует отзывы и реакцию прошлых пользователей, чтобы предлагать продукты другим.

Коллаборативная фильтрация состоит из двух подходов:

Подход на основе пользователей и подхода на основе модели.

User-based.

Подход, основанный на пользователях, предлагает элементы на основе аналогичных профилей взаимодействия с пользователем или аналогичных элементов, с которыми пользователь взаимодействовал.

Item-based.

Подход, основанный на объектах, использует информацию, относящуюся к взаимодействиям пользователя и элементов, для создания скрытой модели для рекомендации.

- Основанные на знаниях

Эти типы рекомендательных систем используются в определенных областях, где история покупок пользователей меньше. Оценки пользователей в данном типе не учитываются. Точность модели оценивается на основе того, насколько полезен рекомендуемый элемент для пользователя.

Самым весомым преимуществом данного типа можно выделить высокую точность рекомендации.

- Гибридные

Гибридные рекомендательные системы – это системы, которые используют как контент, так и совместные данные для предложений. В системе сначала выполняется рекомендация контента, поскольку пользовательские данные отсутствуют, затем после использования системы устанавливаются пользовательские предпочтения с аналогичными пользователями.

Эти системы помогают преодолеть некоторые ограничения рекомендательных систем, такие как проблемы разреженности и холодный запуск.

Редкость данных мешает системе находить похожих пользователей, потому что активные пользователи оценили небольшое количество продуктов или элементов. Проблема холодного запуска возникает, когда системе трудно предоставить точные предложения из-за недостаточной информации о пользователях.

### **Проблемы рекомендательных систем**

Как и любая технология, рекомендательные системы не лишены проблем. Вот самые распространенные из них.

- **Холодный запуск**

Эта проблема возникает, когда в систему добавляются новые пользователи или новые элементы, новый элемент не может рекомендоваться пользователям изначально, когда он вводится в систему рекомендаций без какой-либо оценки или обзоров, и, следовательно, трудно предсказать выбор или интерес пользователей, что приводит к менее точным рекомендациям.

- **Разреженность**

Часто случается, что большинство пользователей не выставляют оценки или обзоры приобретенным товарам или контенту, и, следовательно, модель оценки становится очень разреженной, что может привести к проблемам с разреженностью данных, это снижает возможности поиска группы пользователей с аналогичными оценками или интересами.

- **Синонимия**

Синонимия возникает, когда один элемент представлен двумя или более разными именами или списками элементов, имеющих схожие значения, в таком состоянии система рекомендаций не может распознать, показаны ли термины разные элементы или один и тот же элемент.

- **Конфиденциальность**

Как правило, человеку необходимо передать свою личную информацию (иметь опыт гиперперсонализации) в систему рекомендаций для получения более полезных услуг, но это вызывает проблемы с конфиденциальностью и безопасностью данных, многие пользователи не решаются вводить свои личные данные в системы рекомендаций, которые страдают от проблем с конфиденциальностью данных.

- **Масштабируемость**

Одна из самых больших проблем – это масштабируемость алгоритмов, имеющих реальные наборы данных в системе рекомендаций, огромное количество изменяющихся данных генерируется в результате взаимодействия пользователя с элементом в форме оценок и обзоров, и, следовательно, масштабируемость является большой проблемой для этих наборов данных.

- **Задержка**

Мы наблюдаем, что многие продукты чаще добавляются в базу данных рекомендательных систем, пользователям рекомендуются только уже давно существующие продукты, так как новые добавленные продукты еще не оцениваются.

Таким образом, возникает проблема задержки. Для решения этой проблемы можно использовать метод совместной фильтрации и подход на основе категорий в сочетании с взаимодействием с пользователем и элементом.

### **Оценка качества**

Работая с рекомендательными системами, не стоит забывать об оценке алгоритма. Это нужно, чтобы мы имели возможность сравнивать типы рекомендательных систем для выявления оптимальной системы для той или иной задачи.

- Традиционные метрики
- Добавление порядка в рекомендацию

### **Проектирование рекомендательной системы**

При проектировании нашей рекомендательной системы мы использовали инструменты, наиболее подходящие для этого. Так как задача нашего проекта – это рекомендация фильмов, нам нужна большая база данных с кино, пользователями и их оценками.

Наш проект рекомендательной основан на:

- База данных MovieLens

В качестве источника данных служит базу данных MovieLens, которая находится в общем доступе и является совершенно бесплатной. Данная база данных состоит из шести csv-файлов. Нам нужно из них всего 2 – это “movies” и “ratings”.

- Язык программирования Python и библиотеки Python: Numpy, Pandas, Matplotlib и Seaborn

Выбор языка сделан в сторону Python, так как у данного языка есть очень удобные библиотеки для выполнения нашего проекта. Python сам по себе является отличным языком программирования общего назначения, но с помощью нескольких популярных модулей (numpy, pandas, matplotlib и seaborn) он становится мощной средой для научных вычислений

Работа над проектом будет состоять трех частей:

- Работа с данными

Реализ чтение файлов из БД и выведем нужные нам таблицы.

- Анализ данных

В этой части работы, мы графически анализируем результаты, что получили при работе с данными. Для того, чтобы лучше понимать данные, что мы имеем, построим несколько гистограмм, рассмотрим их и сделаем некоторые выводы.

- Реализация рекомендации

Проделав большую работу с данными и их анализом, можно наконец приступить к созданию самих рекомендаций. Цель нашей работы проект рекомендательной системы, которая бы, по выбранному нами фильму,

подбирала нам похожие фильмы по оценкам, поставленным пользователями конкретно этому фильму. То есть, нам нужно сходство между оценками данного фильма и искомых фильмов у конкретных пользователей.

Результат нашего проекта - рекомендация фильмов, основанная на близости предпочтений пользователей.

### **Применение рекомендательных систем на практике**

Теперь, когда мы рассмотрели все виды, проблемы, преимущества и недостатки рекомендательных систем, мы посмотрим на некоторые крупные компании в арсенале которых имеются самые мощные и продвинутые рекомендательные системы. Ведь только крупные игроки рынка имеют достаточные ресурсы для разработки и применения самых передовых технологий.

На данный момент лучшие примеры рекомендательных систем имеют:

- Amazon
- Spotify
- YouTube
- TikTok

### **Заключение**

Важно понимать, что рекомендательные системы являются развивающейся областью it, поэтому имеют очень хорошую перспективу на будущее. Но даже сейчас они остаются крайне незаменимыми во многих крупных отраслях и компаниях. Рекомендательные системы имеют много весомых и уникальных преимуществ перед другими решениями, но не стоит забывать, что для их реализации требуются хорошие и грамотные специалисты.

В ходе нашей работы мы подробно разобрали понятие рекомендательных систем, их типы и принципы работы, а также их проблемы. Заглянули за кулисы ведущих мировых компаний, чтобы понять, как именно у них получается так хорошо предугадывать желания своих клиентов. Но самое главное, мы спроектировали свою систему рекомендаций, проанализировали данные о

пользователях и фильмах, и получили результат в виде персонализированной рекомендации фильмов на основе близости предпочтений.