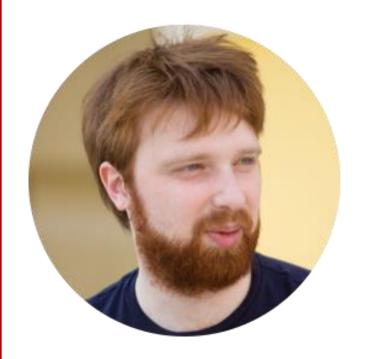
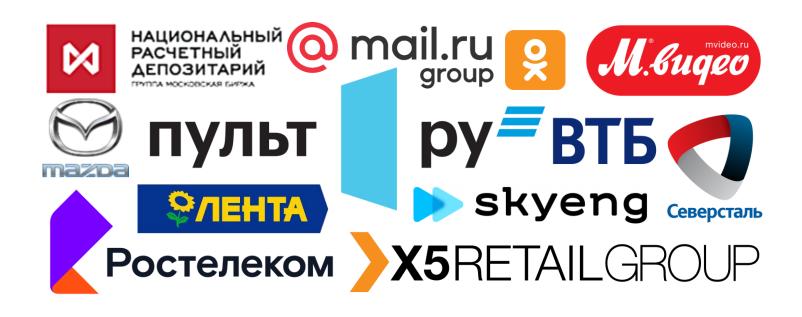
0. О курсе

Чернобровов Алексей









- 1. Введение в рекомендательные системы
- 2. Коллаборативная фильтрация (Collaborative filtering)
- 3. Матричные разложения
- 4. Задачи ранжирования
- 5. Оценка качества рекомендательных систем
- 6. Content-based рекомендации
- 7. Гибридные архитектуры
- 8. Нейронные сети в рекомендациях
- 9. Are We Really Making Much Progress?
- 10. Active Learning в рекомендательных системах
- 11. Высоконагруженные рекомендательные системы
- 12. Внедрение рекомендательных системы

Основная задача курса – дать слушателям представление о том, как строятся рекомендательные системы в реальных проектах.

Домашние задания

Д3 №1

Нужно самостоятельно реализовать колоборативную фильтрацию методом knn, SVD, SVD++ и сравнить результаты работы алгоритмов.

Д3 №2

Реализовать самостоятельно метрики ранжирования и Learning-to-Rank алгоритм.

Д3 №3

Необходимо будет построить нейронную сеть в стиле Neural Collaborative Filtering, обучить ее и сравнить работу с SVD (то есть ДЗ №1 нужно будет все равно сделать).

Д3 №4

Построить гибридную рекомендательную систему (то есть ДЗ №3 нужно будет все равно сделать).

Система оценок

ДЗ №1 – вес 10% (оценка от 0 до 10)
Задача выдается после Зей лекции. Дедлайн 2 недели. ДЗ №2 – вес 20% (оценка от 0 до 10)
Задача выдается после 5ой лекции. Дедлайн 2 недели. ДЗ №3 – вес 20% (оценка от 0 до 10)
Задача выдается после 7ой лекции. Дедлайн 2 недели. ДЗ №4 – вес 40% (оценка от 0 до 10)

Задача выдается после 8ой лекции. Дедлайн 3 недели.

Тест – вес 10% (оценка от 0 до 10) На семинаре после 11ой лекции.

Система оценок

ВАЖНО!

ДЗ через неделю после дедлайна принимается с коэффициентом 0.8. ДЗ через 2 недели после дедлайна принимается с коэффициентом 0.5.

Итоговый бал считается как взвешенное среднее.

Для зачета нужно набрать 4 и больше.

1. Введение в рекомендательные системы

Введение

Разработка качественных рекомендательных систем сегодня является одной из важных целей на пути успеха многих компаний. Эта лекция будет посвящен введению в понятие рекомендательных систем и разбору того, какие виды систем сегодня пользуются наибольшим спросом. А также, в каких областях они применяются и какие методы помогут в их построении.

План

Что такое рекомендательные системы

Виды рекомендаций



Области применения рекомендательных систем

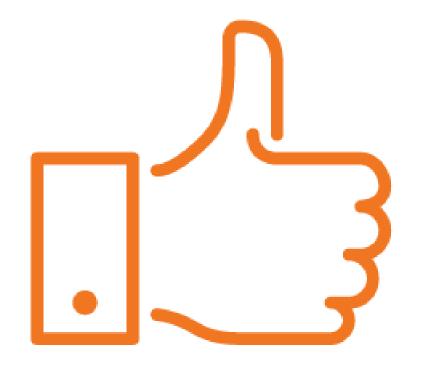
Готовые реализации рекомендательных систем



Кейсы

Что такое рекомендательная система?

Рекомендательные системы — это программы, которые пытаются предсказать, какие объекты (фильмы, музыка, книги, новости, това ры) будут интересны пользователю, на основе определенную информацию о его профиле.























Customers Who Bought This Item Also Bought







Apple iPad MC705LL/A (16GB, Wi-Fi, Black) NEWEST MODEL

**** (360) \$509.95



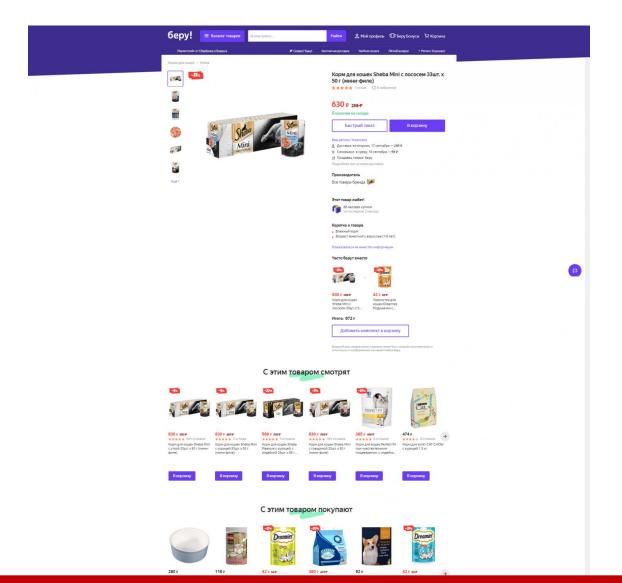
3 Pack of Premium Crystal Clear Screen Protectors for Apple iPad

大大大大大 (1,221) \$1.69

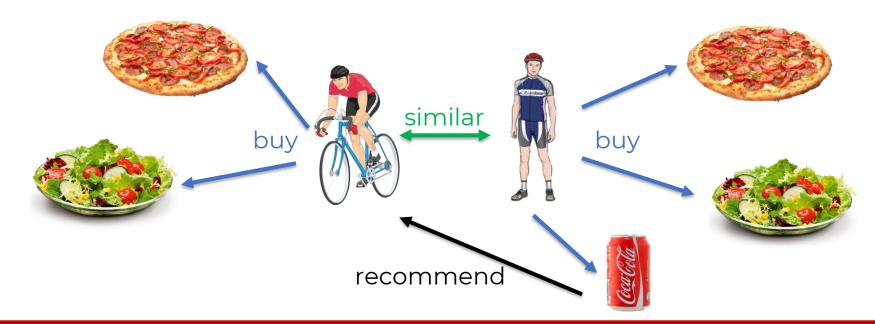


3 Pack of Universal Touch Screen Stylus Pen (Red + Black + Silver)

★★☆☆ (2,082) \$1.10



Такие системы значительно упрощают поиск релевантных продуктов и обогащают опыт пользователя. Множество компаний используют такие платформы для продвижения своих продуктов и услуг, руководясь запросами покупателей. В данном случае рекомендации основываются на историях поиска пользователей.



Релевантность — это мера того насколько хорошо объект (документ, товар) или удовлетворяет потребности пользователя в данный момент.

Например, пользователь, который вводит запрос в поисковую систему ожидает, что результаты будут соответствовать интенту (поисковому намерению) и контексту (времени, месту, погодным условия) запроса.

Какие бывают рекомендации? Где взять данные? Где их показывать?

Примеры рекомендательных систем

Рекомендации:

 \bigcirc

Контента (фильмы, музыка, книги)

Предложение нового контента, повышающего заинтересованность пользователей.



(2)

Товаров

Предложение наиболее интересных товаров в интернетмагазине.



(3)

Событий (концертов, туров)

Предложение наиболее интересных мероприятий для клиента.



Сбор данных

Явный (explicit)

- оценка объекта
- ранжирование группы объектов
- выбор одного товара из двух
- создание списка любимых объектов
- купил

Неявный (implicit)

- что искал, смотрел,
- что клал в корзину
- лог поведения
- анализ содержимого компьютера

Что мы знаем о пользователях

Общая информация:

- Устройство / браузер / размер экрана
- Регион
- Пол
- Дата рождения

Поведенческие факторы (неявный отклик):

- Просмотренные страницы (экраны)
- Время на сайте или в приложении
- Клики

Обратная связь (явный отклик):

- Рейтинги
- Отзывы
- "Лайки«
- Покупки

Виды рекомендуемых товаров

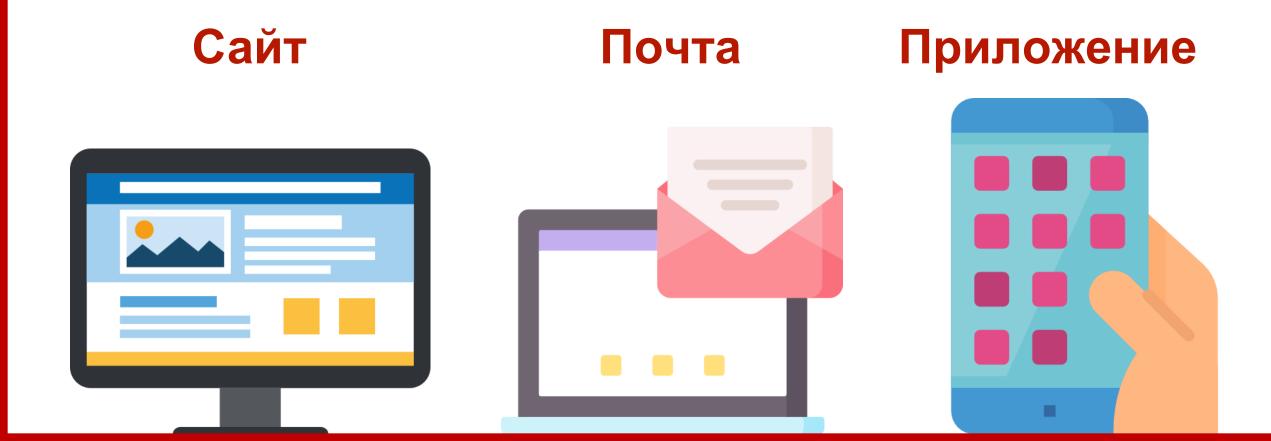
По типу:

- заменители (alternative)
- сопутствующие товары (cross sell)
- бандлы
- более дорогие товары (up sell)
- популярные товары (best sellers)

По способу:

- персональные / неперсональные
- оффлаин / онлаин

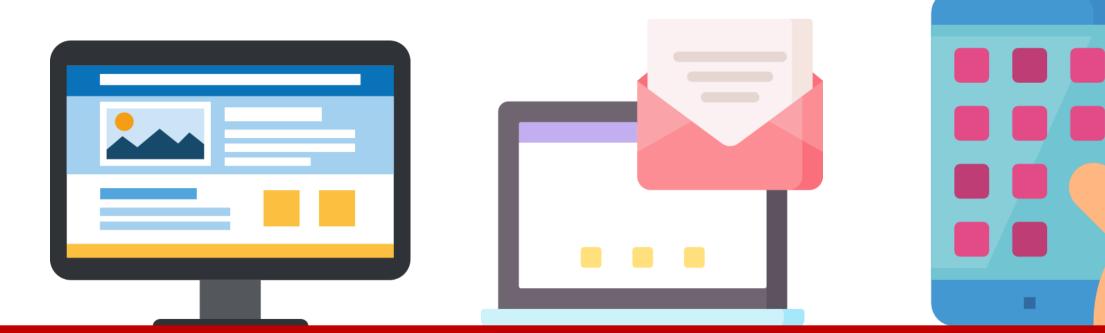
Каналы рекомендаций



Каналы рекомендаций

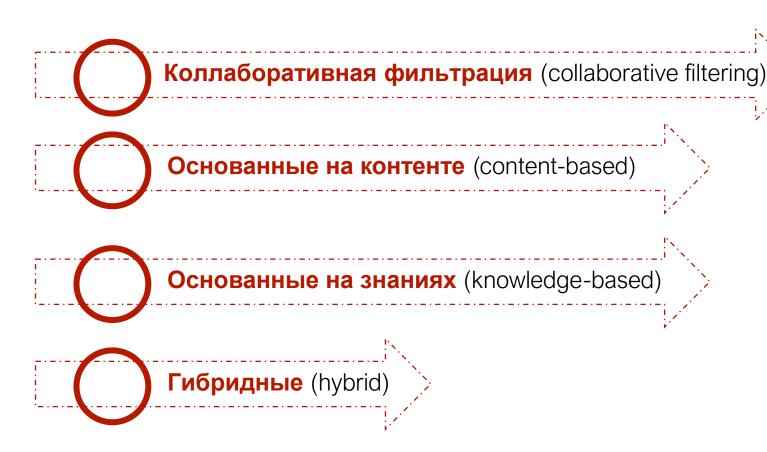
А какие еще каналы можно использовать?

Сайт Почта Приложение



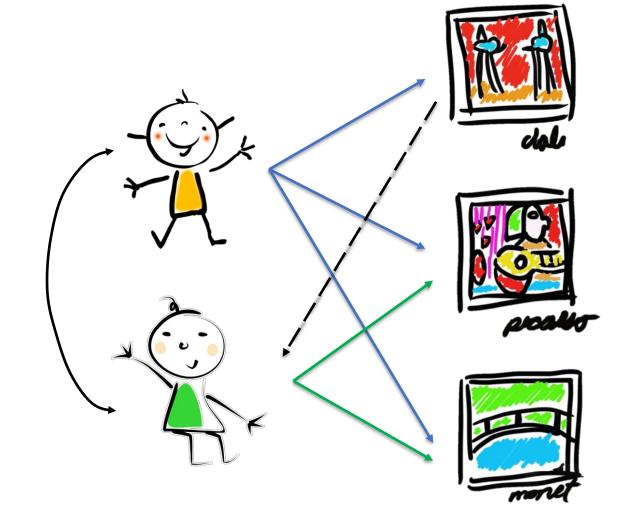
Виды рекомендаций

Виды рекомендаций



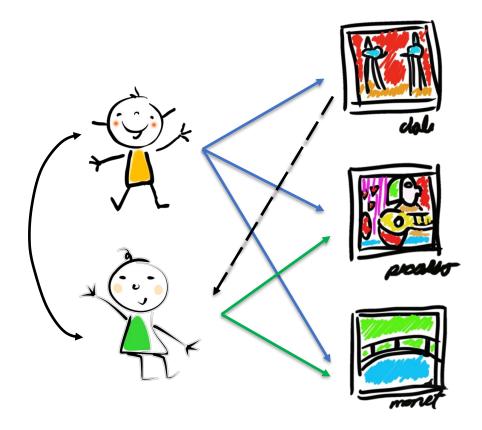
Коллаборативная фильтрация

Рекомендации для пользователя строятся на основе **оценок** похожих пользователей.

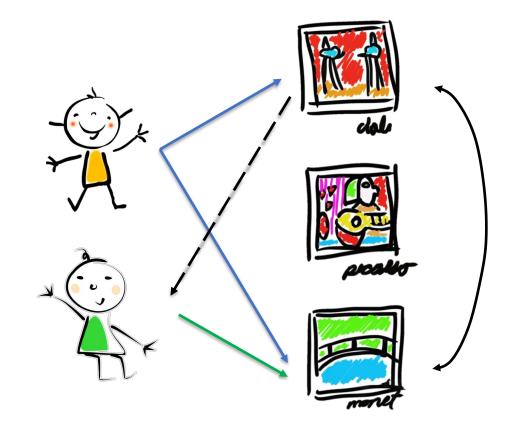


Коллаборативная фильтрация

1 User-based



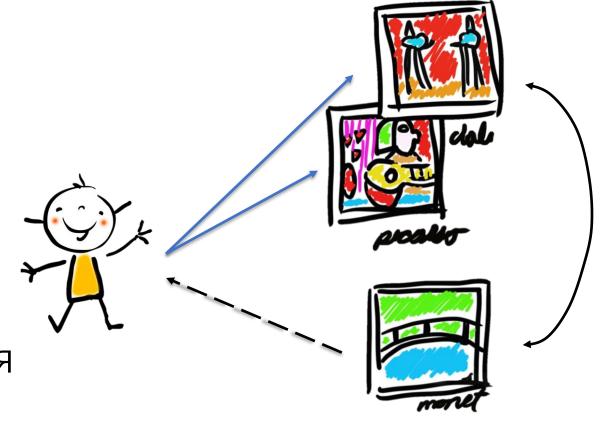
2 Item-based



Основанные на контенте

• Рассчитываются признаки для пользователей и объектов

 Строится модель классификации/ регрессии, приближающая оценки пользователей



Основанные на знаниях

- Строится база знаний о том, как объекты одной предметной области соотносятся с интересами и предпочтениями пользователя, которая с помощью правил эти соотношения описывает.
- Далее, на основе предпочтений пользователей оценивается полезность объектов по этим правилам, и на основании этой полезности строятся рекомендации.



Гибридные

Распространённые типы комбинирования:

- реализация по отдельности коллаборативных и контентных алгоритмов и объединение их предположений
- включение некоторых контентных правил в коллаборативную методику
- включение некоторых коллаборативных правил в контентную методику
- построение общей модели, включающей в себя правила обеих методик

Области применения рекомендательных систем

Netflix Prize

Netflix Prize - это конкурс, в котором требовалось спрогнозировать оценку пользователями фильмотеки Netflix.

Это была задача с явными рейтингами, оценки ставились по шкале от 1 до 5.

Были доступны следующие данные:

- Обучающие данные (training data set) содержат 100.480.507 оценок, которые 480.189 клиентов поставили 17.770 фильмам
- Названия и годы выхода в прокат всех 17.770 фильмов.

Нужно было предсказать какие оценки поставит пользователь тому или иному фильму.



Netflix Prize

Этот конкурс породил бум рекомендательных систем!

Определение победителя:

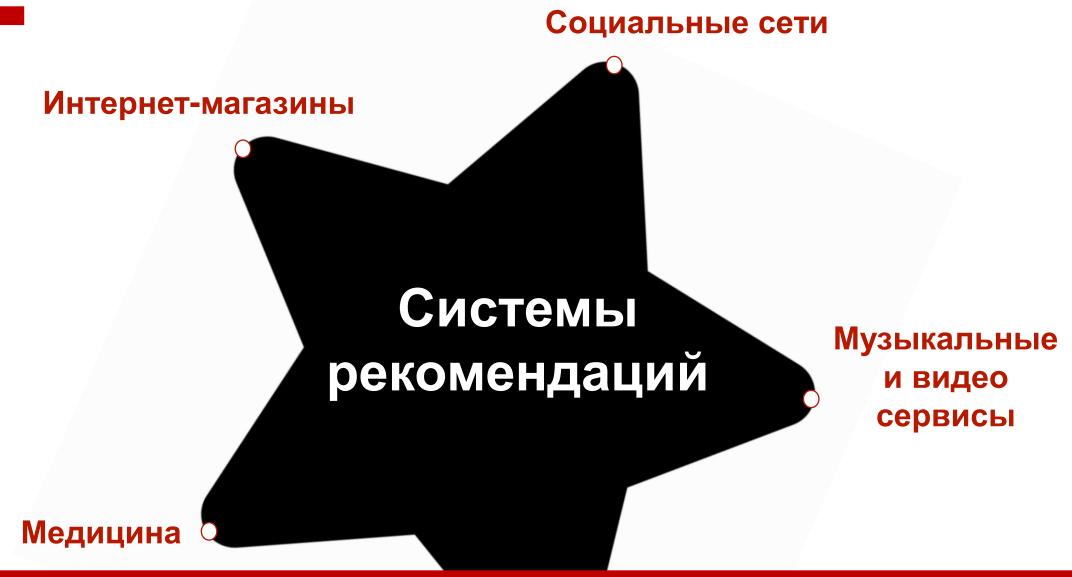
На скрытой части оказалось, что точность у этих команд совпадает до четвертого знака после запятой, поэтому победителя определила разница саббмтов в 20 минут.

Победивший ансамбль использовал модели следующих классов:

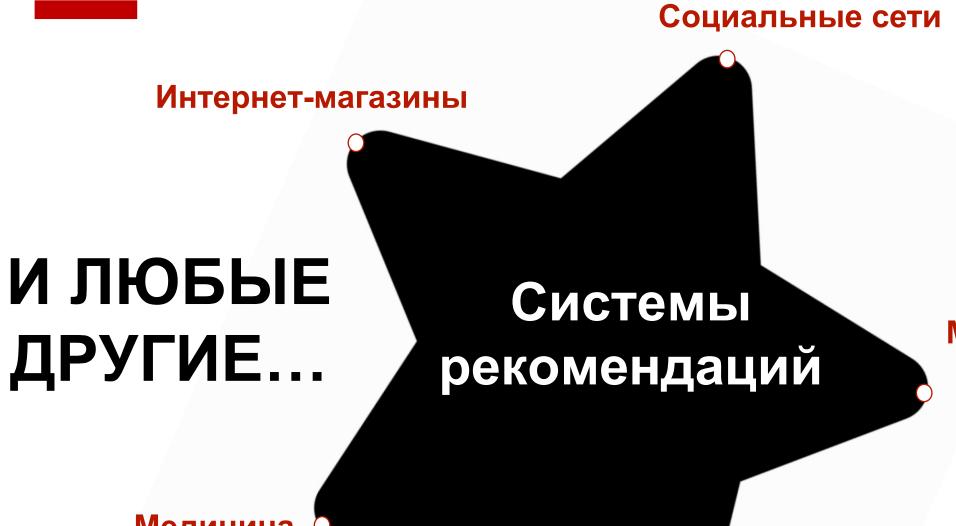
- Регрессионная модель, основанная на средних оценках
- collaborative filtering коллаборативная фильтрация
- Random Forests

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries!	0.8591	9.81	2009-07-10 00:32:20
6	<u>PragmaticTheory</u>	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BiqChaos	0.8601	9.70	2009-05-13 08:14:09
8	<u>Dace</u>	0.8612	9.59	2009-07-24 17:18:43

Области применения



Области применения



Музыкальные и видео сервисы

Медицина

Интернет-магазины

Блок «Похожие товары» в карточке товара

Блок рекомендаций, которые система автоматически сочла подходящими к текущему товару. Таким образом, не нужно указывать связи между товарами, интеллектуальная система сама образует связи товаров между собой.



22 614 P 38 999 P

37 222 P 50 990 P

28 072 ₽ 38-990-₽

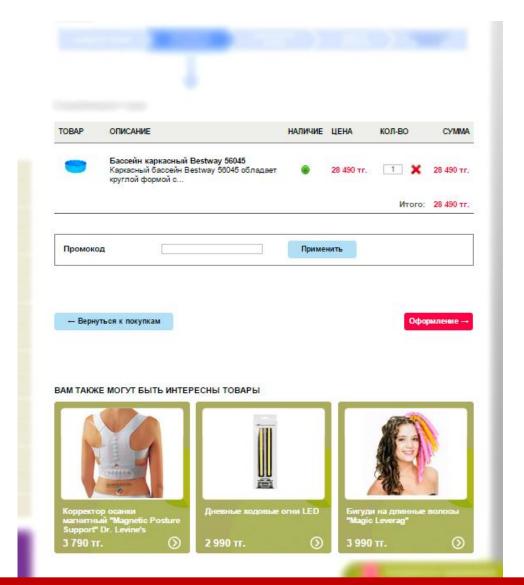
Блок «Популярные товары» на главной странице

Блок популярных товаров, которые чаще всего заказываются в вашем интернет-магазине.



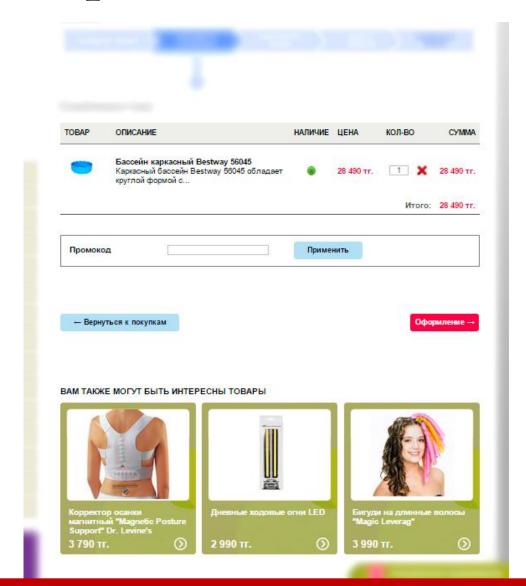
Блок «Экспертные рекомендации» или «Вам также могут быть интересны»

После оформления заказа и перехода в корзину интернет-магазины часто предлагают также к покупке сопутствующие товары, которые могут быть интересны потребителю. Тем самым это стимулирует увеличение среднего чека.



Блок «Экспертные рекомендации» или «Вам также могут быть интересны»

Сотрудники интернет-магазина вручную указывают в административной панели «сопутствующие» для каждого товара. Система фильтрует товары и показывает покупателю те, которые могут ему пригодиться. Таким образом, средний чек увеличивается.



Рекомендательные системы

Лабиринт

Книжные магазины заботливо напоминают нам о том, что отложенная книга поступила в продажу, и предлагают товары, подобранные на основе наших предыдущих покупок.

/абиринт

— Товар появился на складе!



Искусство любить

193 p.

вы хотели заказать этот товар. Рады сообщить, что он снова появился на складе.

Заказать сейчас

Лабиринт рекомендует



Иметь или быть? 202 р.



Бегство от свободы 202 р.



Поступай как женщина, думай как...

272 p.



Маленький принц 184 р.



Анатомия фитнеса и силовых упражнений...

726 p.



Сто лет одиночества **597 р.**

Социальные сети

TikTok

В TikTok лента рекомендаций отражает предпочтения, уникальные для каждого пользователя. Система рекомендует контент путем ранжирования видеороликов на основе набора факторов.

Основные факторы:

- Действия пользователей
- Информация о видео, которая может включать звуки и хэштеги
- Параметры устройства и учетной записи (например, настройки языка, геолокации и тип устройства)



TikTok

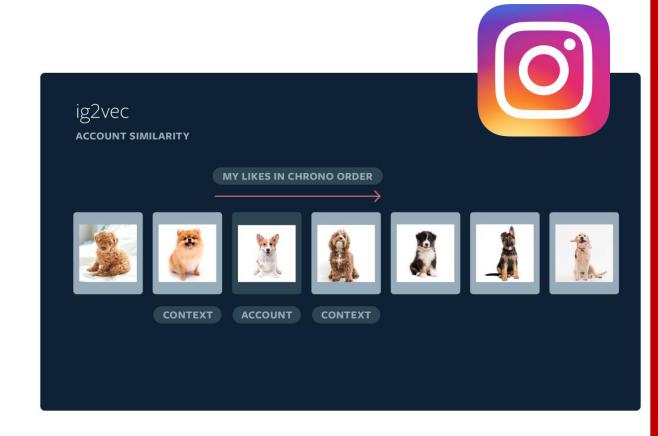
Блоки архитектуры TikTok:

- 1. создание тегов контента
- 2. создание пользовательских профилей и историй (сценариев, scenarios)
- 3. обучение и обслуживание алгоритмов рекомендаций



Instagram

Основной фокус в алгоритме направлен на анализ учетных записей пользователей в Instagram. Для анализа используется **ig2vec** — фреймворк наподобие word2vec. Метод позволяет эффективно анализировать семантику языка и строить векторно-семантические модели на основе контекстной близости слов в тексте.



Instagram

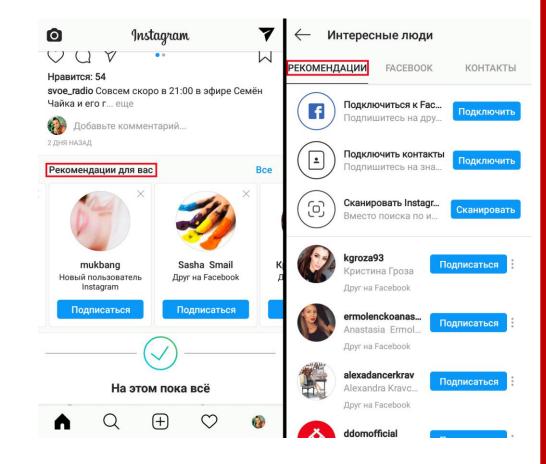
2 этапа работы рекомендательной системы:

1. Поиск источников

Алгоритм, используя данные о взаимодействиях пользователя (лайки, сохранения, комментарии, расшаривание) собирает стартовый набор данных

2. Фильтрация контента

В дальнейшем происходит фильтрация для исключения нерелевантного контента. На выходе получается порядка 500 вариантов релевантного контента, которые в дальнейшем отправляется на этап ранжирования.

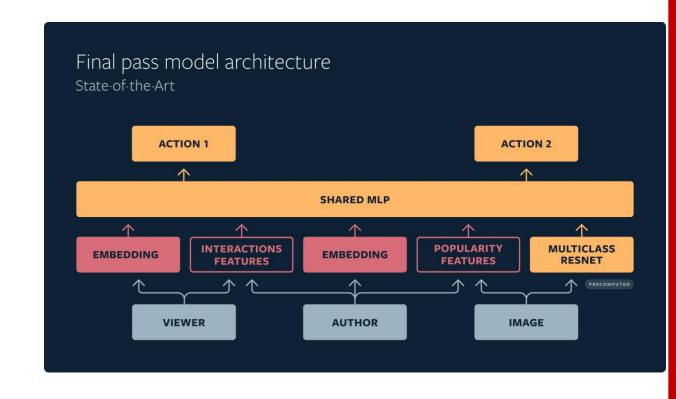


Instagram

На входе алгоритм имеет 500 подходящих вариантов, которые нужно ранжировать.

Этот процесс делится на 3 итерации:

- Первая: происходит отбор 150 лучших вариантов из 500.
- **Вторая:** упрощенная нейросеть с полным набором функций отбирает 50 самых качественных вариантов.
- Заключительная: более сложная нейросеть с полным набором функций отбирает 25 лучших и наиболее подходящих вариантов для первой страницы рекомендаций.



Музыкальные и видео сервисы

Киносервис imdb.com

Сайт IMDb.com — крупнейшая в мире база данных кинофильмов, сериалов и телешоу.

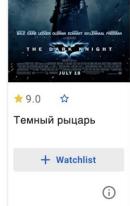
Под каждым фильмом можно найти раздел «More like this», что значит, что эти фильмы рекомендуют вам на основании схожести ваших вкусов со вкусом других людей. Очевидно, что в данном сервисе используется коллаборативная фильтрация.

More like this



+ Watchlist

(1)





Бойцовский клуб

Watch options

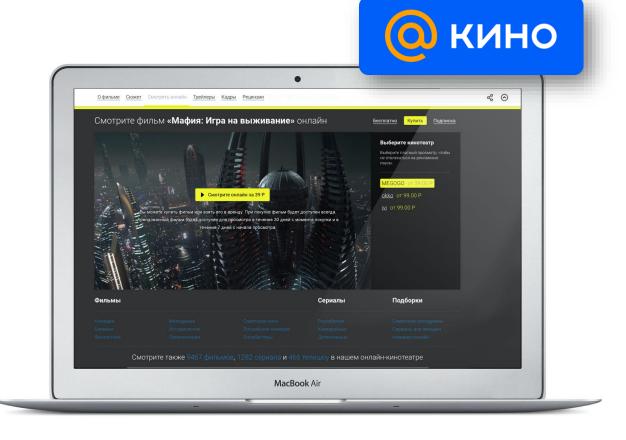
(i)



Кино Mail.ru

В проекте «Кино Mail.Ru» используется несколько моделей рекомендаций. Их можно сгруппировать по решаемым ими задачам:

- Генерирование «холодных» рекомендаций для пользователей с малым количеством оценок (и даже совсем без них).
- «Горячие» модели для пользователей, активно ставящих оценки. Пользователь становится «горячим» после того, как поставит десять оценок.



Музыкальные сервисы



В основе работы музыкальных сервисов (Apple.Music, Spotify, Яндекс.Музыка) лежит коллаборативная фильтрация.



Однако, есть две проблемы:



- Холодные пользователи это те, кто только что зарегистрировался в сервисе, и система о них ничего не знает
- **Холодные треки это** те, которые только появились, и о них ничего неизвестно, либо какие-то редкие песни.

Музыкальные сервисы

Решение проблемы с холодными пользователями:

Новичками после регистрации предлагают рассказать, кого они слушают. После этого система понимает, что можно предложить конкретному пользователю. А если пользователь не хочет отвечать на вопросы, ему накидывается средняя популярная музыка, и система смотрит на реакцию. Если пользователь переключает трек, алгоритм понимает — не зашло, и подбрасывает что-то из другого жанра.

Музыкальные сервисы

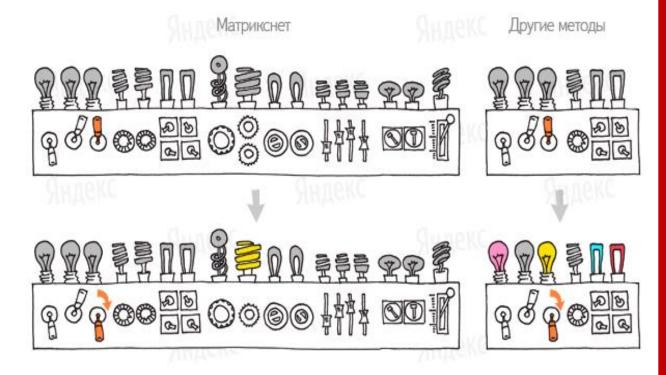
Решение проблемы с холодными треками:

- Алгоритмы в буквальном смысле гуляют по сети, анализируют тексты и таким образом видят, что пишут об исполнителях и об их треках.
- Сервисы анализируют каждый загруженный в систему аудиофайл. Трек нарезается на фрагменты по ним делаются спектрограммы, понятные машинам. Так нейросети обучают на «внутренностях» популярных треков, по которым много отзывов. После этого алгоритм подбирает что-то похожее по звучанию из тех холодных треков, которые мало кто слышал.

Яндекс.Музыка

- «Яндекс.Музыка»
 самостоятельно строит
 плейлиты на основе
 предпочтений пользователя, а
 также его поведения
- Список рекомендаций составляется с помощью «Матрикснета» разработанного в «Яндексе» метода машинного обучения





Медицина

Рек. системы в здравоохранении

Основание для рекомендаций — это заполненная пользователем информация о физическом состоянии, хронических заболеваниях, занесенный в программу рацион питания и активность человека в течение дня.

Человек в этом случае получает рекомендации на основе личной статистики и общепризнанных медицинских рекомендаций по поддержанию здоровья.

Рек. системы в здравоохранении

Примеры использования:

• Постановка диагноза

Имея медицинскую карту пациента, методами коллаборативной фильтрации можно показать, какие заболевания чаще всего встречаются у людей, имеющих похожую историю болезни, и, таким образом, предсказать заболевания, которые могут быть пока не диагностированы у пациента.

• Советы пациентам

В особенности для тех, нуждается в постоянном приеме лекарств и отслеживании показателей здоровья. Больные сахарным диабетом могут заносить данные о чувстве голода, болях, уровне глюкозы и приеме лекарств в мобильное приложение, которое агрегирует все показатели и напоминает о приеме необходимых лекарств.

Рек. системы в здравоохранении

Примеры использования:

• Оптимизация больничных закупок

Построение рекомендаций о том, какие препараты нужно добавить в ассортимент, чтобы повысить удовлетворенность пациентов и избежать дефицита необходимых лекарств.

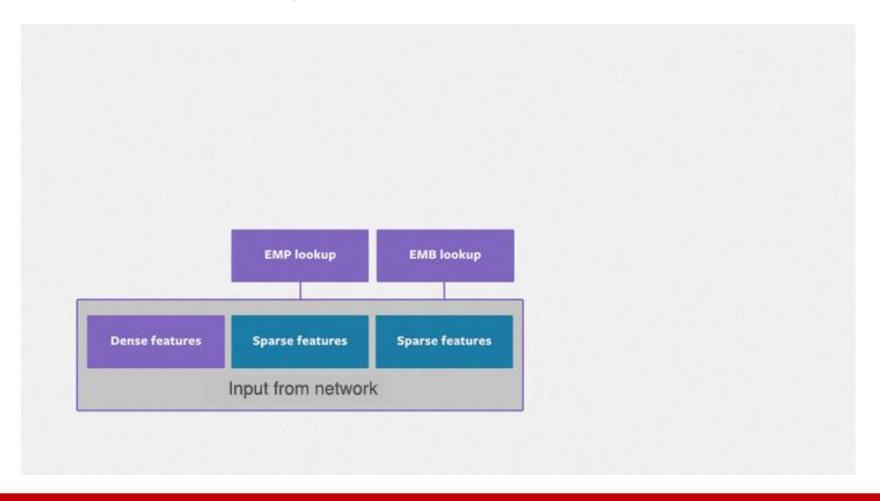
• Предоставление рекомендаций врачам

Врачи через рекомендации могут получить ссылки на похожие случаи заболевания, описания сопутствующих заболеваний и симптомов, ссылки на протоколы лечения и литературу, информацию о том, какие лекарственные средства дают наибольший эффект при лечении конкретного заболевания.

Готовые реализации рекомендательных систем

Библиотека DLRM от Facebook

Библиотека DLRM на Python и Caffe2 для построение рекомендательных систем.



Библиотека DLRM от Facebook

Принцип работы:

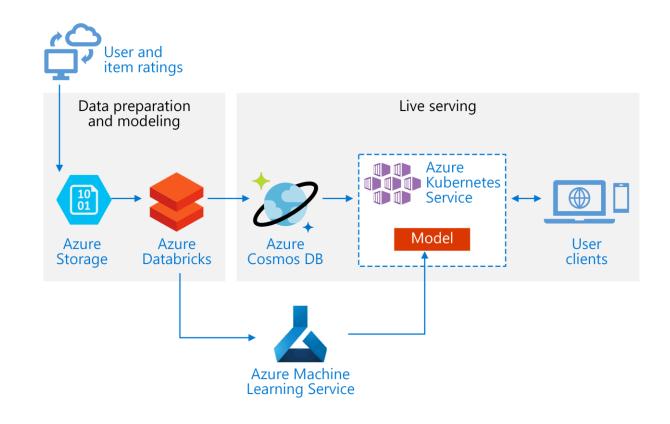
В DLRM категориальные характеристики обрабатываются с использованием векторных представлений. В свою очередь, численные характеристики проходят через многослойный перцептрон (MLP). Затем напрямую вычисляются второстепенные взаимодействия характеристик. В итоге, результаты обрабатываются в многослойном перцептроне и передаются в функцию сигмоиды, чтобы получить вероятность клика для товаров.

Рекомендательная система обрабатывает численные (плотные) и категориальные (разреженные) данные, которые описывают пользователей и продукты.

Recommenders of Microsoft

https://github.com/microsoft/recommenders

- Десятки разных алгоритмов, которые постоянно пополняются
- Встроенные dataset'ы
- Удобство интеграции с Azure



TensorFlow Recommenders

https://www.tensorflow.org/recommenders

TensorFlow Recommenders (TFRS) – это библиотека для построения рекомендательных системных моделей.

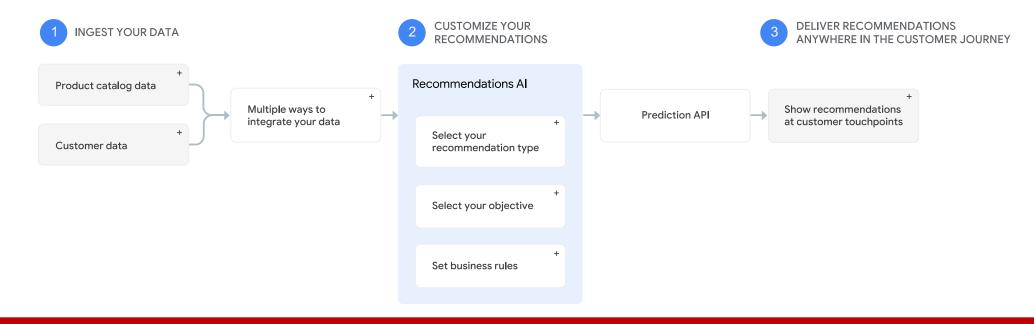
- Помогает в полном рабочем процессе построения рекомендательной системы: подготовка данных, формулирование модели, обучение, оценка и развертывание.
- Построен на Keras и нацелен на плавное обучение, но при этом дает вам гибкость для создания сложных моделей.



Recommendations AI от Google AI

https://cloud.google.com/recommendations

Рекомендации AI позволяет создать непрерывную персонализированную систему рекомендаций, основанную на современных моделях машинного обучения с глубоким обучением, без необходимости иметь опыт в машинном обучении или системах рекомендаций.



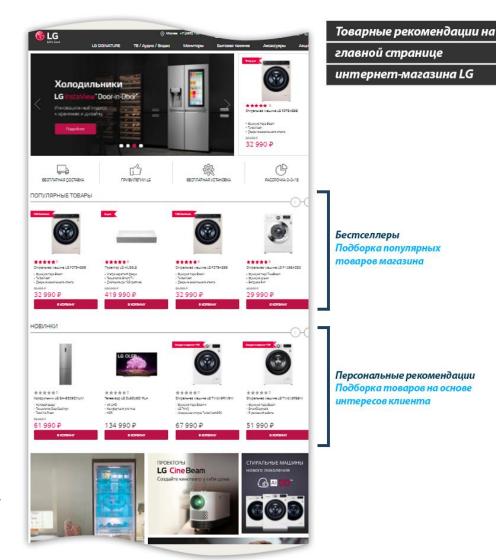
Кейсы

Интернет-магазин LG

Кейс:

Разработка комплексной стратегии персонализации товарных предложений в интернет-магазине и запуск автоматизированной email-кампании, а также повышение эффективности триггерных сценариев с помощью технологии Growth Hacking'a.

- 10% выручки обеспечивают рекомендации
- 13% заказов содержат рекомендованные товары
- 13% выручки от триггерной коммуникации генерируют письма после совершения заказа

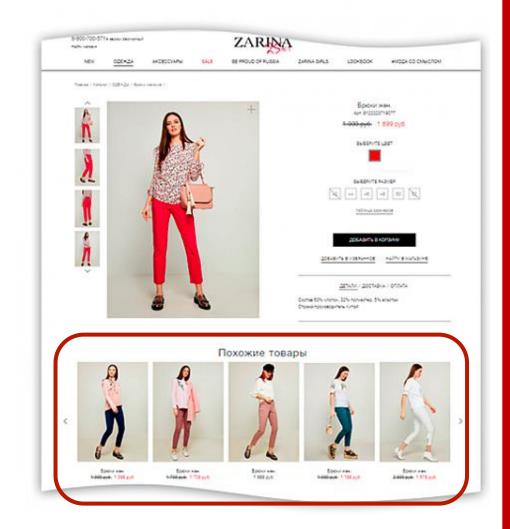


Бренд одежды Zarina

Кейс:

Внедрение рекомендательной системы на различных страницах сайта: в карточке товара «похожие товары», в корзине «с этим товаром часто покупают», на странице категории «хиты категории»

- увеличение конверсии на 12,85%
- рост выручки на 10,41%



Интернет-магазин Рив Гош

Кейс:

Для более 220 торговых точек интернет-магазина «Рив Гош» в 2019 году была разработана обучаемая рекомендательная система, предсказывающая поведение покупателей.

- увеличение лояльности клиентов
- увеличение продаж
- снижение издержек на маркетинговые акции



IKEA



Кейс:

Использование сервиса Recommendations AI от Google для более точной настройки рекомендательной системы, настройка различных конфигураций: «Рекомендовано для вас», «Часто покупаются вместе» и «Другие, которые могут вам понравиться».

- IKEA Retail (Ingka Group) увеличила среднюю стоимость заказа в целом по миру для на 2% с помощью Recommendations AI
- увеличение количества релевантных рекомендаций на 400%
- улучшение частоты кликов на 30%

Заключение

- Улучшение ленты на главной странице автоматически увеличивает покупки с нее на 30%
- Меньше 10% смотрят на ленту главной страницы
- Только качественно разработанная и персонализированная рекомендательная система может увеличить выручку компании и другие показатели