

Рекомендательные системы
для бизнеса и не только

Matrix Factorization (Part 1)

This is a two-part, two-week module on matrix factorization recommender techniques. It includes an assignment and quiz (both due in the second week), and an honors assignment (also due in the second week). Please pace yourself carefully -- it will be difficult to finish in two weeks unless you start the assignments during the first week.



Video · Introduction to Matrix Factorization and Dimensionality Reduction



Video · Singular Value Decomposition



Video · Gradient Descent Techniques



Reading · On Folding-In with Gradient Descent



Video · Deriving FunkSVD



Video · Probabilistic Matrix Factorization

- Алгоритмы для рекомендаций хорошо известны
- Оптимизируют одну метрику, которую не так сложно взломать
- Зачем нужны большие команды разработчиков для рекомендательных систем?

Знакомство с рекомендательными системами

Постановка задачи

- Даны множество пользователей и множество объектов
- Для некоторых пар (пользователь, объект) известен результат взаимодействия
- Требуется предсказывать результат для всех пар (пользователь, объект)
- Обычно на основе этих предсказаний для пользователя отбираются N самых релевантных объектов

Отбор кандидатов

- Объектов очень много
- Нужно сначала сузить множество поиска для пользователя
- Используются быстрые методы
- Примеры: user-based CF, kNN, SVD

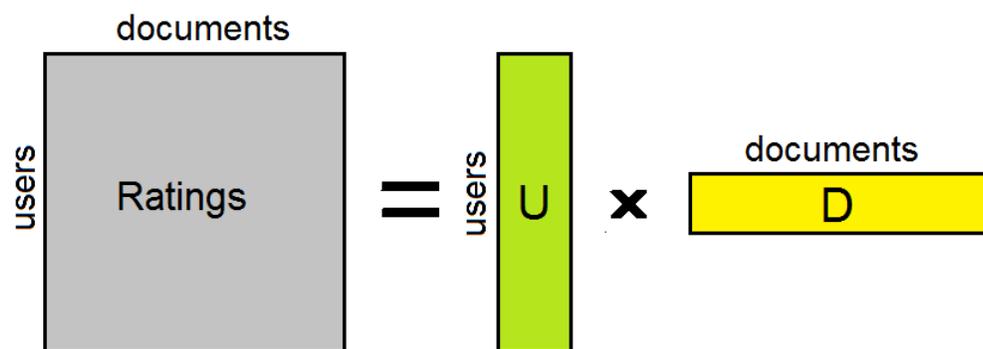
Ранжирование

- После отбора кандидатов нужно тщательно выбрать лучших
- Вычисляются факторы для пар (пользователь, объект)
- Используется тяжёлая модель (композиция деревьев, нейросеть) для предсказания результата взаимодействия

Рекомендации в Дзене

Отбор кандидатов в два этапа:

- Отбор источников/авторов
- Отбор документов с источников



Рекомендации в Дзене

Ранжирование:

- CatBoost
- Целевая переменная: длинные клики, время
- Несколько сотен факторов (от простых статистик до нейросетевых моделей)

Рекомендации в Дзене

Статистические факторы ранжирования

- Неперсональные (CTR/дочтения документа и домена, явный фидбек)
- Персональные (то же самое, но для данного пользователя)

Рекомендации в Дзене

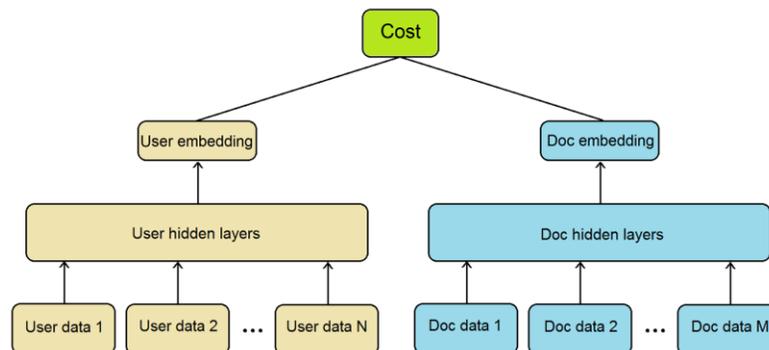
Документные факторы ранжирования

- Возраст
- Тематики
- Контентные классификаторы
- Векторные представления статьи

Рекомендации в Дзене

Парные факторы ранжирования — похожесть на документы из истории пользователя

- Разные документы из истории: клики, лайки, дизлайки и т.д.
- Разные векторные представления
- Разные способы измерения сходства: DSSM, матричные разложения и т.д.



Ранжирование в Дзене

Реранкеры:

- Разнообразие
- Учёт подписок
- Свежесть
- ...

Где всё это используют?

История про target



Интересные предложения

РЕКОМЕНДУЕМ



Кулеры и системы охлаждения д...
Deepcool NEPTWIN V2

от 2 548 ₽

ХИТ ПРОДАЖ



Защитные пленки и стекла для т...
Защитное стекло Xiaomi

от 149 ₽

ТОВАР НЕДЕЛИ · **ДО -20%**



Электрообогреватели и тепловы...
Ballu BOH/CL-07

от 1 630 ₽

СКИДКИ **ДО -10%**



Цифровые бытовые метеостанции
Oregon Scientific LW301

от 8 990 ₽

РЕКОМЕНДУЕМ



Жесткие диски, SSD и сетевые н...
Samsung MZ-7KE512BW

Территория Grohe



ХИТ ПРОДАЖ



Дрели, шуруповерты, гайковёрты
Интерскол ДА-12ЭР-01 ...

РЕКОМЕНДУЕМ



Жесткие диски, SSD и сетевые н...
Samsung MZ-7KE1T0BW

Похожие товары



Canon EOS 200D Kit

от 32 390 ₽

1 отзыв 109 предложений

Любительская зеркальная фотокамера

Байонет Canon EF/EF-S

Объектив в комплекте, модель

уточняйте у продавца

Матрица 25.8 МП (APS-C)

Цвет:

Цены 109



4.0

Canon EOS 750D Kit

от 32 650 ₽

5 отзывов 133 предложения

Любительская зеркальная фотокамера

Байонет Canon EF/EF-S

Объектив в комплекте, модель

уточняйте у продавца

Матрица 24.7 МП (APS-C)

Цвет:

Цены 133



Canon EOS M10 Kit

от 21 250 ₽

4 отзыва 89 предложений

Фотокамера с поддержкой сменных объективов

Байонет Canon EF-M

Объектив в комплекте, модель

уточняйте у продавца

Матрица 18.5 МП (APS-C)

Цвет:

Цены 89



до -30%



Canon EOS M6 Kit

от 38 300 ₽

до 84 940 ₽

2 отзыва 105 предложений

Фотокамера с поддержкой сменных объективов

Байонет Canon EF-M

Объектив в комплекте, модель

уточняйте у продавца

Матрица 25.8 МП (APS-C)

Цвет:

Цены 105



С ЭТИМ ТОВАРОМ ЧАСТО ПОКУПАЮТ



4.0

HIPER MP15000

Универсальные внешние аккумулято...

от 1 453 ₽

30 отзывов 33 предложения

Цвет: ●

Цены 33



2.5

Apple EarPods (Lightning)

Наушники и Bluetooth-гарнитуры

от 1 590 ₽

6 отзывов 66 предложений

Цвет: ○

Цены 66



Сетевая зарядка Apple

Зарядные устройства и адаптеры

от 642 ₽

16 предложений

Производитель: Apple

Тип: сетевая зарядка

Разъем подключения: for Apple (Lightning)

Цены 16



Smart cover Apple

Чехлы для планшетов

от 3 080 ₽

4 предложения

Производитель: Apple

Совместимость: Apple

Тип: smart cover

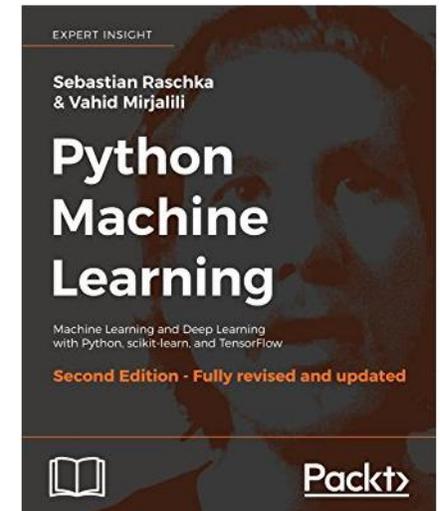
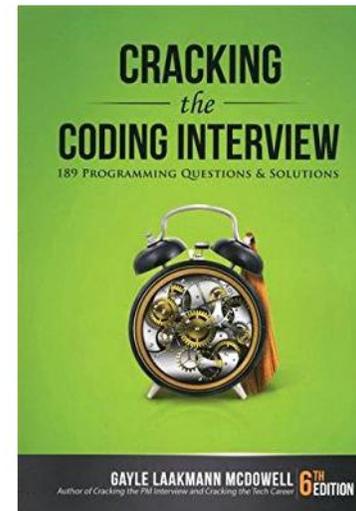
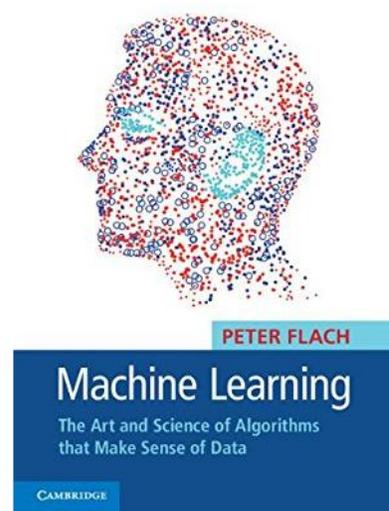
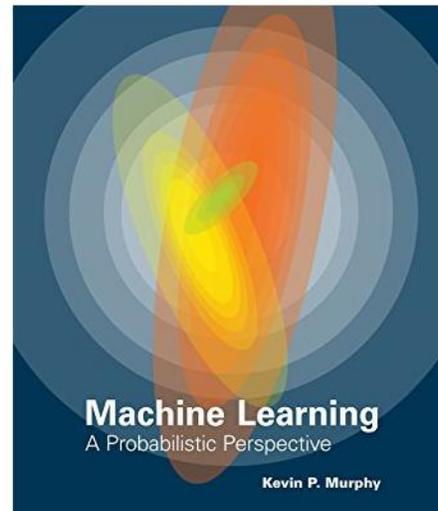
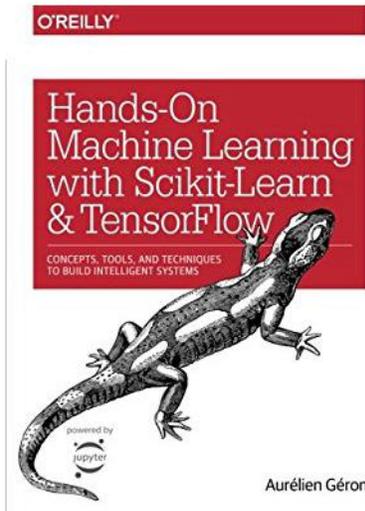
Функция подставки: Да

Цвет: ● ●

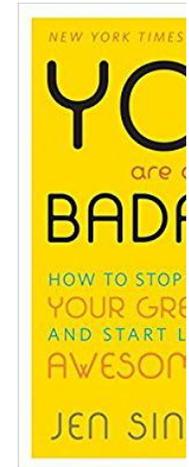
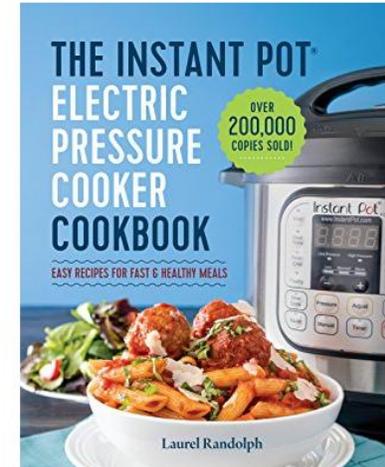
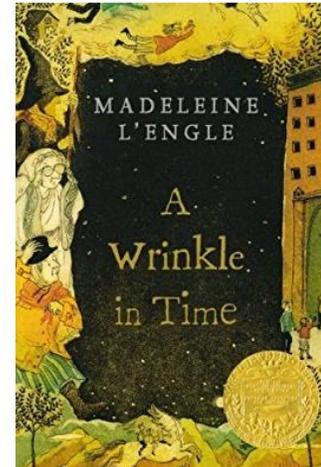
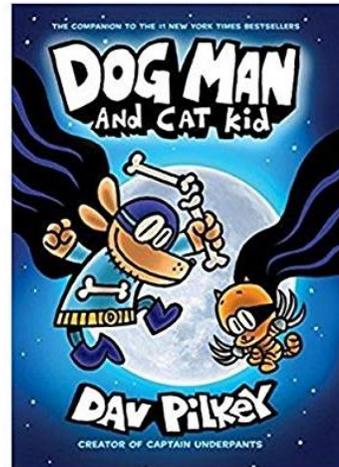
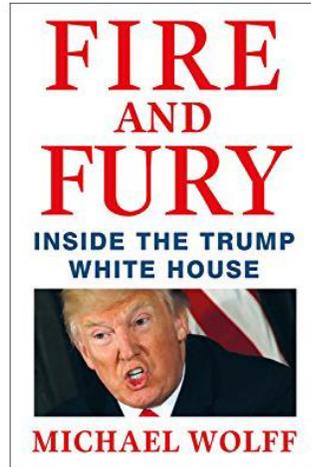
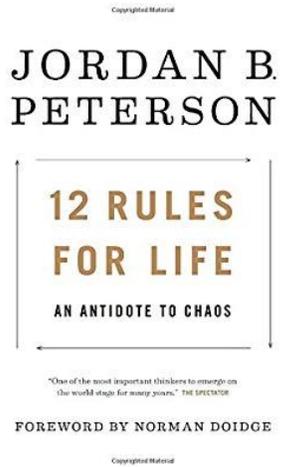
Цены 4



Recommendations for you in Books



Books best sellers [See more](#)



Рекомендации в онлайн-ритейле

- Персональные предложения для каждого пользователя на основе большого числа факторов
- Учёт статистики по всем пользователям сервиса
- Новый взгляд на маркетинг
- Работа маркетолога — правильное использование рекомендательных механизмов, выбор каналов для взаимодействия ИИ и пользователей
- Amazon — 35% заказов из полок рекомендаций

Рекомендации в маркетинге

- Персональные коммуникации вместо сегментации
 - Почтовые рассылки, SMS, звонки в магазинах
 - Up-sell в банках
- Удержание клиентов и повышение лояльности
 - Поиск клиентов с большой вероятностью ухода, выбор оптимальной стратегии удержания

NETFLIX

Browse ▾

Kids

DVD

EPISODES

Because you watched New Girl



TV Mysteries







Что мешает городским железным дорогам

Железная дорога – это не только
рельсы и шпалы для перевозки...

VARLAMOVS.RU



13 сюрпризов неайтишной компании

Отработав много лет в
компаниях, занимающихся
разработкой программного
обеспечения на заказ, невольно
начинаешь задумываться — а как
всё-таки выглядит вся эта суета с...

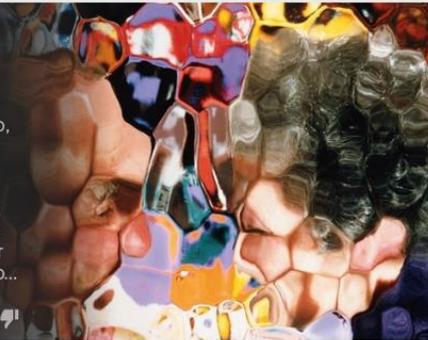
ХАБРАХАБР



Почему настоящие лидеры должны уметь думать медленно

Есть известная (но, возможно,
апокрифическая) история о
том, как теория
относительности пришла в
голову Альберту Эйнштейну,
когда он ехал на велосипеде.
Говорят, что Уоррен Баффетт
читает не меньше шести часо...

HBR-RUSSIA.RU



Математики обыграли букмекеров

Букмекеров можно обыграть,
если учесть, что они иногда
выставляют не самые
выгодные для себя
коэффициенты, чтобы снизить
риски. Международная груп...

N+1



Рекомендации контента

- Медийный бум приводит к взрывному росту объёмов информации в сети
- Рекомендательные системы помогают ориентироваться
- Для авторов — поиск целевой аудитории

- Пионеры в Китае — Toutiao (более 100 миллионов активных пользователей) и другие платформы
- Netflix — 2/3 видео найдены через рекомендации
- Youtube — 60% просмотров благодаря рекомендациям

Кейс 1: up-sell в банке

Рекомендации для кросс-продаж

- Крупный банк с миллионами клиентов
- Кредиты, депозиты, мобильный банкинг, страховка
- Обычные e-mail/sms рассылки
- Используют сегментацию
- При добавлении персонализации -> +13% дохода

Рекомендации для кросс-продаж

- Получили большой массив данных, много факторов
- Обученная модель имеет высокое качество на тестовых данных

Рекомендации для кросс-продаж

- Получили большой массив данных, много факторов
- Обученная модель имеет высокое качество на тестовых данных
- Проанализировали важность факторов, получили странные результаты
- В данных случилось много утечек при склеивании таблиц

Рекомендации для кросс-продаж

- Добились хорошего качества в оффлайне и хороших факторов в топе по важности
- Сравниваем со старой версией в А/В-тесте
- Что может пойти не так?

Рекомендации для кросс-продаж

- Одной группе SMS отправили в пятницу, другой — в субботу
- В группах использовали разные схемы фильтрации каналов
- В ходе обсуждения с заказчиком изобрели 5 способов оценить эффект от рекомендаций
- По 4 из 5 способов победили

Кейс 2: развитие экосистемы

Целевая переменная в рекомендациях

- Успешное взаимодействие пользователя и объекта: клик, длинный клик, просмотр, покупка, прослушивание и т.д.
- Что может пойти не так?

К Земле приближается астероид-убийца - столкновение произойдет через несколько дней

Истерия вокруг Pokemon Go дает свои «плоды»: жизнь собак кардинально изменилась

Савченко сделала шокирующее заявление о переходе на сторону России

Жизнь после смерти существует – шокирующее открытие ученых



Лучшие
ПРИКОЛЫ
Рунета #124



Когда ты болеешь, но на учебу идти надо



СЮДА ИДИ!



ЭТО ТЫ СПАЛ НА
НАШЕЙ КРОВАТИ?



ЗАПРЕТ СЛАДОСТЕЙ!!!ГДЕ КУПИТЬ ВКУСНЯШКИ ?!

Daniel Boy • 290 тыс. просмотров • 9 часов назад

DANIEL BOY ВЛОГ: <https://www.youtube.com/channel/UCzr5s1tVxE9n-6lZCx2fuMQ>



Челлендж КВЕСТ нашла КОРОБКУ с игрушкой My little Pony Cutie Mark Crew НОВАЯ коллекция Пони

My little Nastya ✓ 105 тыс. просмотров • 8 часов назад

Челлендж КВЕСТ Quest challenge Как мы искали КОРОБКУ с игрушкой My little Pony Cutie Mark Crew НОВАЯ коллекция Пони МИЛАШКИ для детей For kids Children <https://youtu.be/ba8-UrO80PQ>



Радужная еда для единорога – 9 идей

Трум Трум ✓ 1,1 млн просмотров • 1 день назад

Подписывайся на канал: <https://goo.gl/kLx8qK> Гигантский антистресс костюм / Как стать Барби: https://youtu.be/PCxkAFtZ53s?list=PLw_MSrc8v85fwwaXVeX10CQzrD3fSWvJR Единороги тоже



ЧТО ЗДЕСЬ НЕ ТАК?

НАСКОЛЬКО РАЗВИТ ТВОЙ МОЗГ? 9 загадок на сообразительность

MOGOL TV ✓ 215 тыс. просмотров • 1 день назад

В этом видео на Моголь ТВ ты сможешь проверить насколько развит твой мозг. Тебя ждут 9 загадок на сообразительность. В комментариях не забудь поделиться своими результатами.

Проблема одной целевой переменной

- Авторы начинают оптимизировать контент
- Возникают кликбейт, дублирование и т.д.
- Выгоднее писать на популярные тематики
- Авторам, которые делают уникальный и тематический контент, сложно выделиться



Компьютерные игры в СССР



Все, кто думает, что советские люди были совсем лишены компьютерных игр, ошибаются. Ведь в конце 80-х в стране получают небольшое распространение компьютеры Spectrum, Atari и другие.



Давайте же
посмотрим, во что
играли первые
поколения советских
геймеров. :)

3D Deathchase

Гонка на мотоцикле по лесу. Нужно было уклоняться от деревьев и атаковать противников.





**Прорыв трубы
сильно испортил
воздух в
Стокгольме**



[http://justnu.ru/
novosti-
stokgolma/878-
proryv-truby-
silno-isportil-
vozdukh-v-
stokgolme](http://justnu.ru/novosti-stokgolma/878-proryv-truby-silno-isportil-vozdukh-v-stokgolme)

Из-за этого
рестораны и
отели, которые
находятся на
улице
Drottninggatan
остались без
посетителей.

Неприятная
ситуация, с
которой
столкнулись

Супернарративы

- Модераторы отсматривают нарративы и помечают классные
- Позже был обучен классификатор для автоматизации отбора
- Классные нарративы добавляются в отдельную ленту и выделяются при ранжировании

Нирвана

- В программу могут попасть только авторы и сайты, соответствующие определённым требованиям
 - Нет кликбейта и спама
 - Контент уникальный
 - Нет навязчивой рекламы
 - и т.д.
- Авторы из программы получают приоритет в ранжировании и особую поддержку

Кейс 3: управление лентой

Алгоритмические рекомендации

- Проблема холодных пользователей
- Проблема необычных пользователей
- Вероятность ошибки всегда есть
- Пользователям нужны инструменты управления лентой

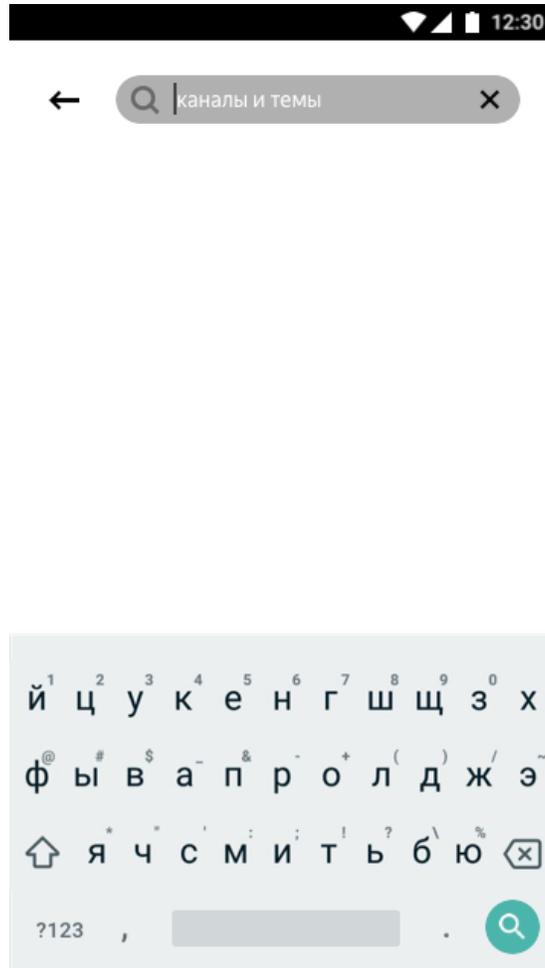
Вариант 1: явный фидбек

- Явные сигналы могут противоречить кликам
- Пример: сайт с некачественной мобильной версией
- Модель по умолчанию не учитывает факторы по лайкам/дизлайкам/блокировкам, нет ощущения управляемости

Вариант 1: явный фидбек

- Вариант 1: большие веса для кликов/некликов, по которым имеет явный фидбек
- Вариант 2: переход к вещественной целевой переменной
- Вариант 3: отбор кандидатов только на основе контентных факторов и явного фидбека
- Вариант 4: отдельная модель для каждого типа фидбека
- Как принимать такие эксперименты?

Вариант 2: темы



Вариант 2: темы

- Как собирать данные?
- Как оценивать качество классификатора?
- Как выстраивать иерархию тем и учитывать её в модели?

Резюме

- Рекомендации — не только алгоритмы для предсказания поведения пользователя
- Оценка качества, развитие тематического контента, управляемость рекомендаций и т.д. — важная часть работы