

Немного о себе

6+ лет опыта работы в Data Science



Васильев
Алексей

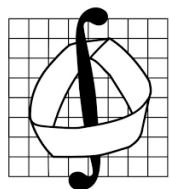
Опыт работы

Несколько лет работал в стартапе
системным аналитиком и руководителем
проектов

Проекты

- Создание библиотеки
рекомендательных систем
RePlay на Spark
- Реком. система для книжного
интернет магазина
- Улучшение качества поиска
- Рекомендации похожих
запросов
- Написание научных статей
- И многое другое

Образование



Выпускник мех-мата МГУ
Аспирантура МГУ



КонсультантПлюс

Руководитель направления
анализа данных



Закончил ШАД Яндекса



Тимлид группы
рекомендательных систем
в AI Lab

Рекомендательные системы повсюду



Главная

Навигатор

Shorts

Подписки

Библиотека

История

Вы сможете ставить

Пикабу

35 минут назад

Вспоминаем старые фильмы, может что-то пропустили или стоит пересмотреть?

Лучшие фантастические фильмы 1992 года: pikabu.ru/link/9ppUThk6EK

Лучшие фантастические фильмы 1992 года

Поделились 32 человека

43 32

Фильм "КАРАТЕЛЬ" / Пробуждающий 2018 | каждой...

ПАПАЙСКАЯ ЧЕРНАЯ ИКРА И ЖЕЛЕЙНЫЙ МЕДВЕДЬ ВАЛЕРА

Познаватель

37 млн просмотров • 4 года назад

ДО ПОСЛЕ

1 You're The First, The Last, My Everything

Barry White

2 Химия

ANNA ASTI

3 We Are Family (1995 Remaster)

Sister Sledge

4 Adore You

Harry Styles

5 Если в сердце живет любовь (Саундтрек к сериалу "Моя любимая Стра...

Mary Gu

6 Love Is In The Air

John Paul Young

Введите запрос

Все Музыка Видеоигры Сейчас в эфире Ситкомы Туристические направления Мультфильмы Кулинария Экшн

Delivery Club

Адрес доставки

Скидка 250 Р от 1300 Р

BOOM

В ресторанах скидка

25

Доставка за 1 Р

FREED

Горячо и сытно

KFC

Рестораны

Магазины

Акции

↑↓ Любимые Навынос Акции Кухни Гастромакет Саларис Акции

THE МЯСО

45-55 мин От 1 Р 4,7

Грузинские Каникулы

55-65 мин От 1 Р 4,5

ПРОШАШЛЫК

40-50 мин От 1 Р 4,5

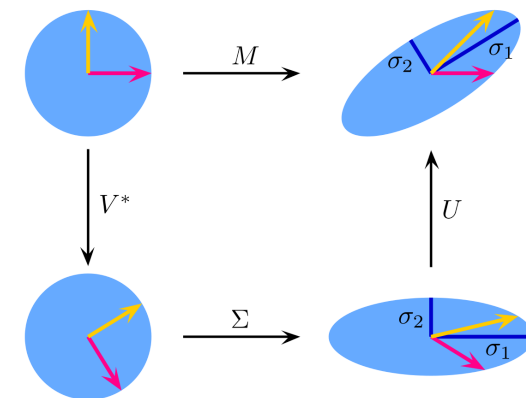
Необходимые навыки для DS в RecSys

Минимум:

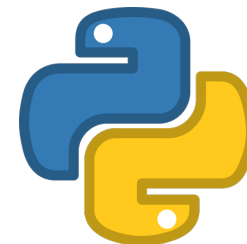
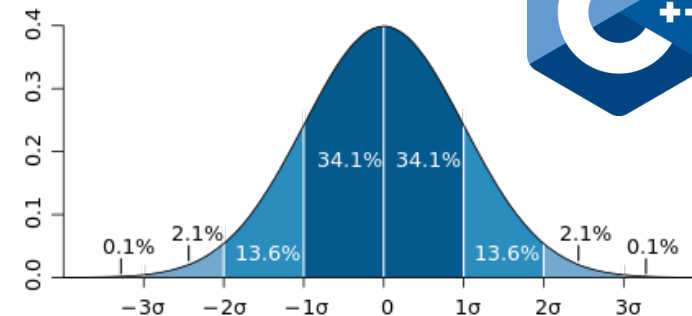
- Линейная алгебра
- Математическая статистика
- Python
- Базовый SQL

Полезно уметь:

- Опыт работы с BigData (Spark, Hadoop)
- Создание микросервисов (Flask)
- Знание компилируемого ЯП (Java, Scala, C++)
- Основы Product Management



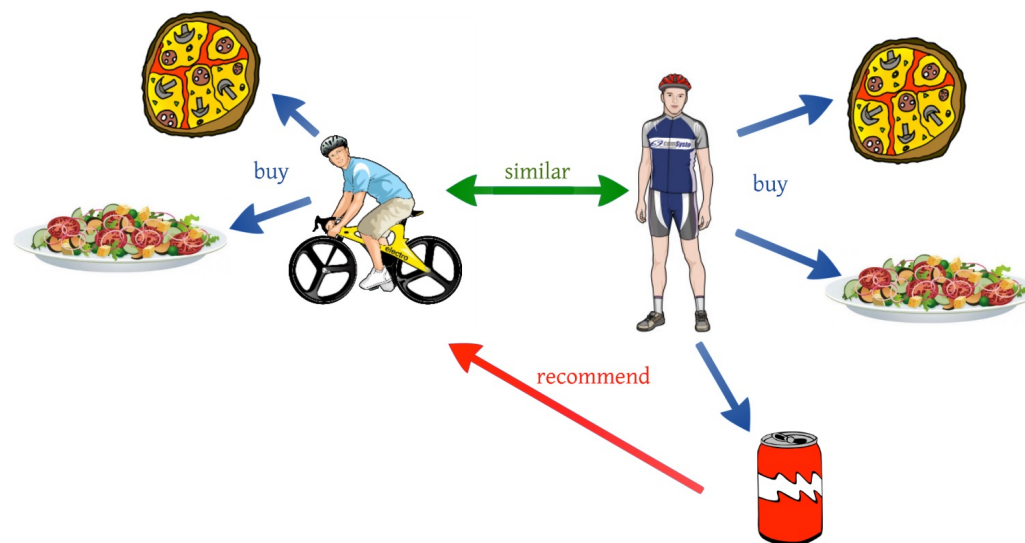
$$M = U \cdot \Sigma \cdot V^*$$



В чем отличие рекомендаций от других задач



- Много пользователей (user) тысячи или миллионы
- Много товаров/объектов (item) тысячи или миллионы
- Взаимодействий значительно меньше чем $\#user * \#items$
- Требуется предсказать несколько объектов
- Пользователи не могут просмотреть все товары
- Могут присутствовать только положительные взаимодействия



Примеры рекомендательных задач



Задача

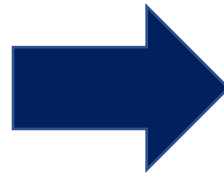
Рекомендации объектов пользователям

Рекомендации пользователей объектам

Рекомендации объектов объектам

Рекомендации пользователей пользователям

Продолжение сессий пользователей



Пример

Предложение фильмов пользователям

Предложения отправить рекламу клиентам

Показать похожие товары на текущий

Пользователи, которых вы можете знать

Прослушивание музыкального контента

Типы отклика в задачах рекомендаций



Explicit

Явный отклик пользователя

- Лайк/дизлайк
- Рейтинг
- Добавление в любимое



Implicit

Неявный отклик пользователя

- Просмотр
- Покупка
- Репост
- Комментарий
- Добавление в корзину
- Загрузка
- Любые действия с объектом



Проблемы при построение рек. систем



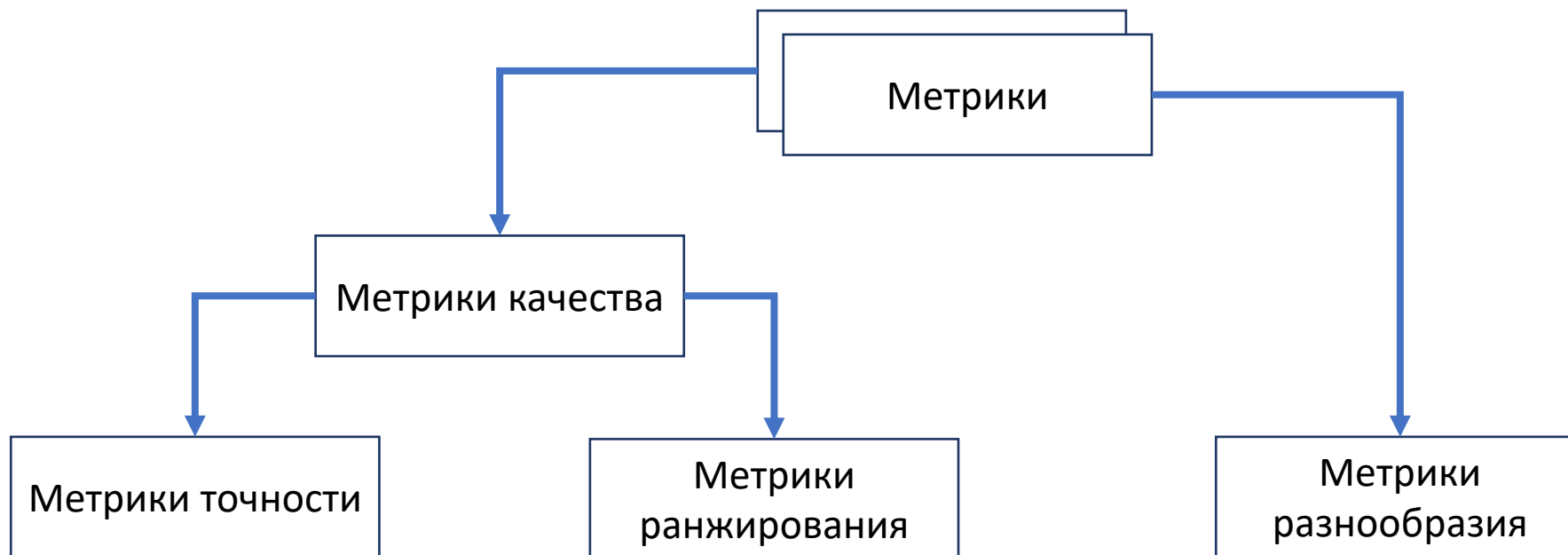
Данных значительно больше, чем в задачах другого типа

«Холодный» старт – новые пользователи и объекты

Попадание в «информационный пузырь»



Как оценивать рекомендательные системы



HitRate
Precision
Recall

Roc Auc
NDCG
MAP
MRR

Coverage
Surprisal
Unexpectedness
Novelty
Serendipity
Diversity



С метриками могут быть проблемы



Library	HitRate@20	MAP@20	MRR@20	NDCG@20	Precision@20	Recall@20	RocAuc
RePlay	0.475	0.039	0.186	0.093	0.058	0.096	0.283
DL RS Evaluation	1.15	0.039	0.186	0.08	0.058	0.096	0.283
DaisyRec	0.475	0.023*	0.275	0.093	0.058	0.096	fail
Beta RecSys	—	0.032	—	0.093	0.058	0.096	0.687
RecBole	0.475	0.039	0.186	0.093	0.058	0.096	0.687
Elliot	0.475	0.073	0.186	0.09	0.058	0.096	0.688
OpenRec	—	—	—	0.463	0.058	0.096	0.705
MS Recommenders	—	0.032	—	0.093	0.058	0.096	0.687
NeuRec	0.475*	0.003	0.186	0.093	0.058	0.096	—
rs_metrics	0.475	0.032	0.186	0.093	0.058	0.096	—

Table 1: Metrics calculated using different libraries.

* The calculation resulted in an error. If the required fix is minimal, we provide results with this mark.

Статья на конференции RecSys'21

Quality Metrics in Recommender Systems: Do We Calculate Metrics Consistently?

<https://arxiv.org/abs/2206.12858>

Library	SAUC	GAUC	GAUC@20	LAUC@20
RePlay	—	—	0.283	—
DL RS Evaluation	—	—	0.283	—
DaisyRec	—	—	fail	—
Beta RecSys	0.687	—	—	—
RecBole	0.687	0.858	—	—
Elliot	0.688	0.705	—	0.112
OpenRec	—	0.705	—	—
MS Recommenders	0.687	—	—	—

Table 2: AUC values distributed by the meaning of the calculated metric.