

РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ



Понятие рекомендательной системы

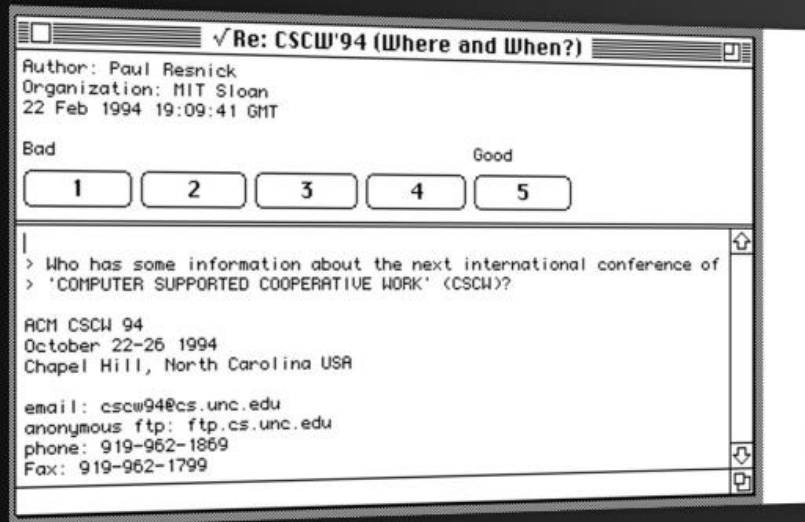
- Рекомендательные системы анализируют предпочтения пользователей и предсказывают, что может понравиться им в будущем.
- Их алгоритмы часто построены на основе машинного обучения: ИИ учится на выборе пользователя и предлагает ему все новые возможности взаимодействия.
- Если пространство предполагает выбор, ему наверняка нужна система для построения рекомендаций.

Системы поддержки принятия решений и рекомендательные системы

- **Рекомендательная система** = выбор в каталоге, помощь в выборе среди большого множества, выбор обусловлен предпочтениями (вкусами) пользователя
- **Система поддержки принятия решений** = рекомендовать лучший путь в логистике, рекомендовать долю вещества в сплаве, рекомендовать какие акции купить или продать ...

Первая рекомендательная система

- Рекомендация научных статей – середина 90-х годов

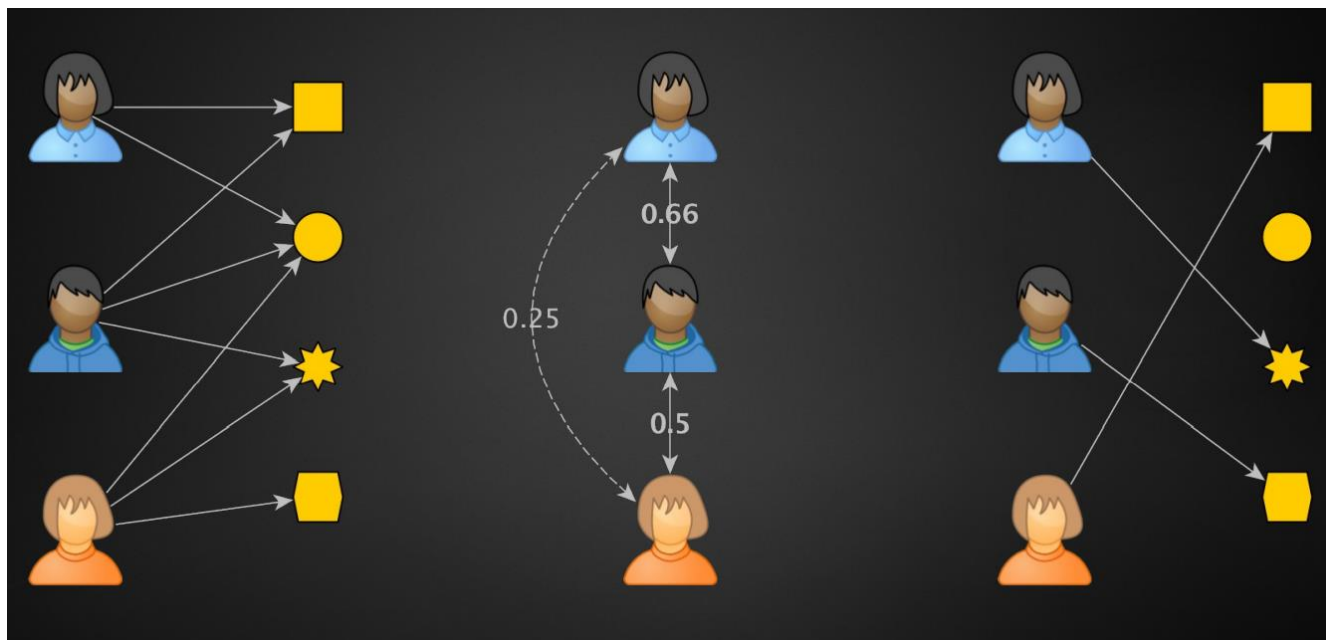


comp.groupware			
▷ 2	██████	P. Flewrat	INFO NEEDED on Groupware 94 Itinerary
▽ 2	██████	Susan McDaniel	awareness information in distributed groupware applications
	██████	Christoph Burkhard	Re: awareness information in distributed groupware applications
▷ 3	██████	Carol Anne Ogden	Re: Lotus Notes for UNIX?
-	██████	Wolfgang Prinz, I3.	CSCW-Workshop: Betrieblicher Einsatz von CSCW-Systemen
▽ 3	██████	Dan Beaton	Scheduling Algorithms
	██████	David Newman	Re: Scheduling Algorithms
	██████	Pete Bergstrom	Re: Scheduling Algorithms

<http://ccs.mit.edu/papers/CCSWP165.html>

Принцип работы рекомендательных сетей

**Вам понравится то, что нравилось тем,
Кому нравилось то, что нравилось вам**



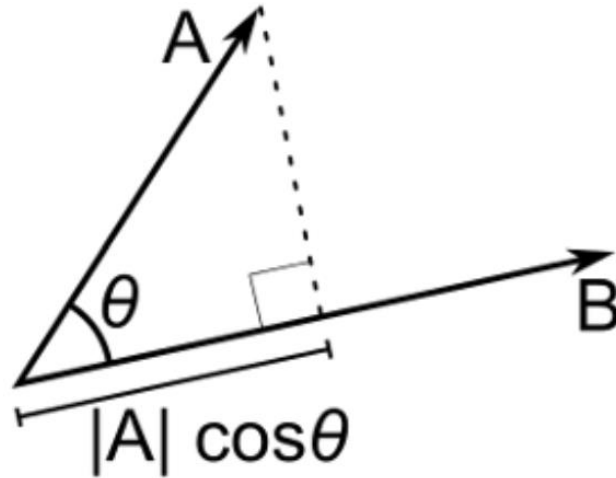
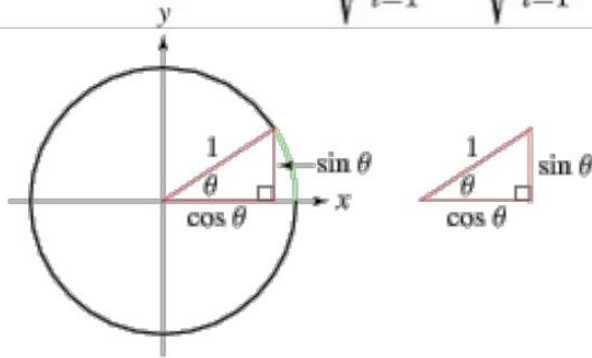
Похожие пользователи

Похожие предметы

Коллаборативная фильтрация: поведение одного пользователя,
влияет на то, что рекомендуется другим пользователям

Метрика похожести на основе косинуса

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

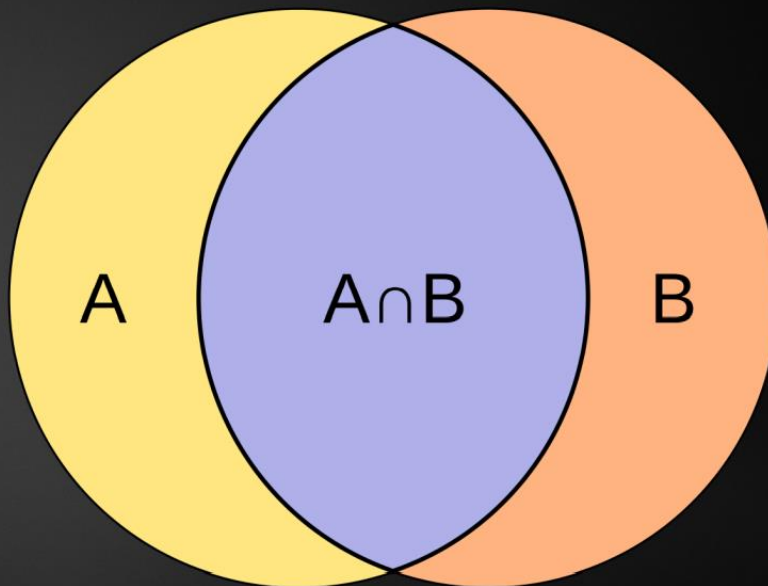


<http://www.ahmetyildirim.com.tr/2018/03/simple-cosine-similarity-for-php.html>

- Оценки – это вектора в многомерном пространстве
- Размерность пространства – количество пользователей (предметов), которые (которым) ставят оценки
- Два пользователя похожи, если их вектора смотрят в одном и том же направлении
- Cos близок к единице вектора сонаправленны

Метрика похожести Жаккара

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$



<https://neo4j.com/docs/graph-data-science/current/alpha-algorithms/jaccard/>
https://en.wikipedia.org/wiki/Jaccard_index

- Множество предметов, которые понравились одному пользователю и множество предметов, которые понравились другому пользователю
- Если результат отношения близок к 1, то множество предметов совпадают

Критерии выбора

**Контентная
фильтрация**
(content-based
filtering)
основывается
на потребленно
м контенте.

Платформа аудиотреков

Похожесть
предметов

Небольшой
стабильный
каталог

Меняющиеся
вкусы клиентов

**Похожесть пользователей или
похожесть предметов**

Новостной портал,
социальная сеть

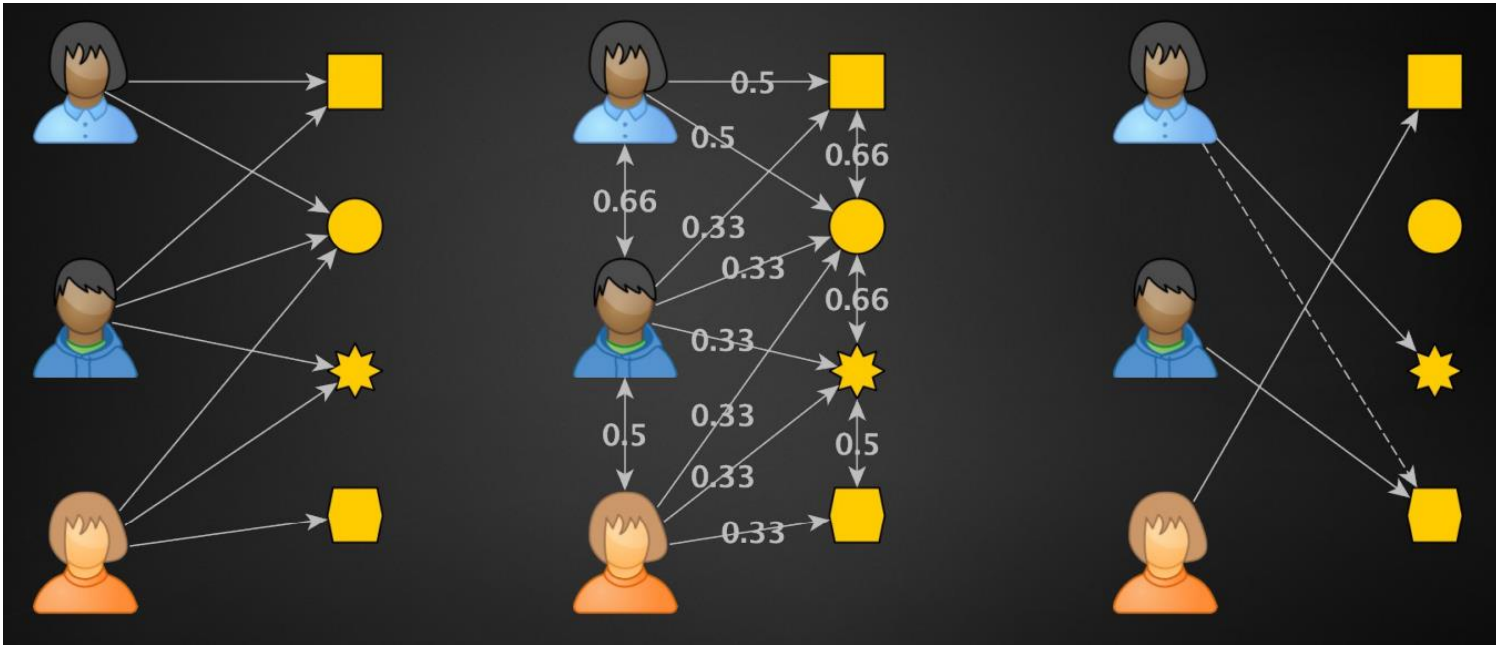
Похожесть
пользователей

Большой
динамический
каталог

Относительно
стабильные
клиенты

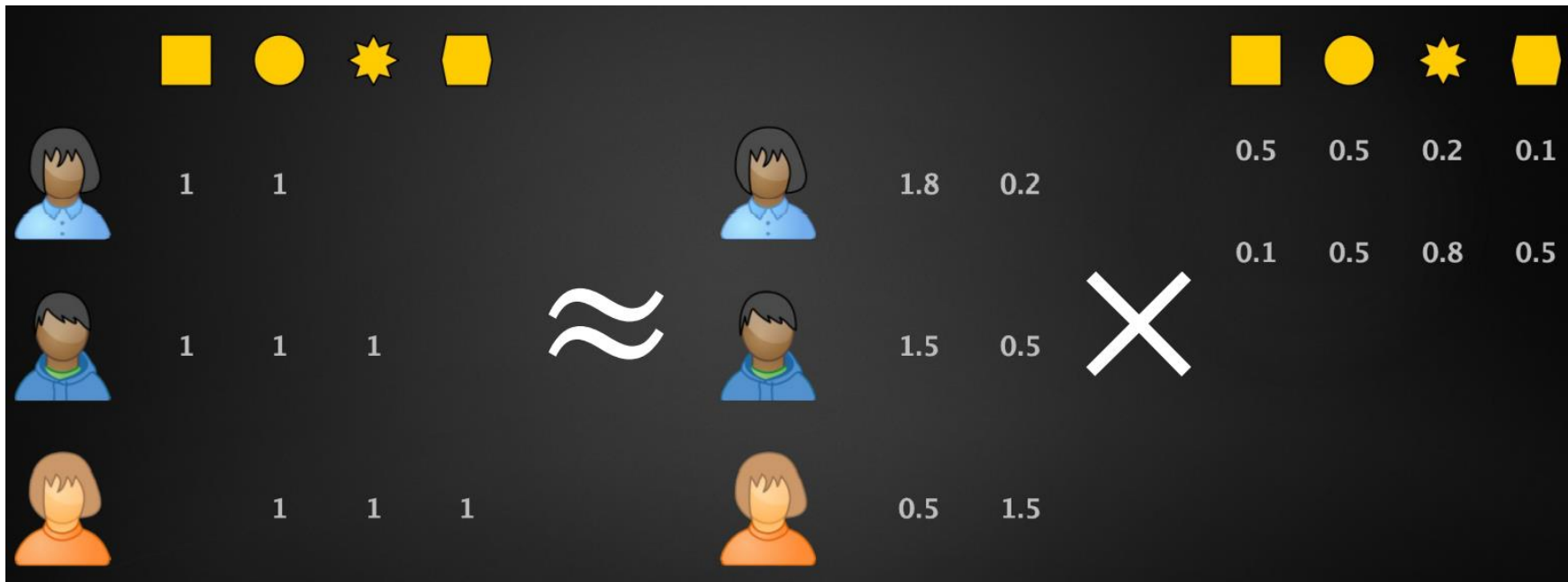
**Коллаборативная
фильтрация** (collaborative
filtering) учится на опыте
других пользователей
со схожими интересами.

Интеграция = похожесть пользователей + похожесть предметов



- Композитный граф похожести
- Движение по графу от конкретного пользователя
- Процесс случайного блуждания по графу
- Вероятность выбора дуги графа пропорциональна весу дуги
- Вероятность возвращения в исходную точку (для персонализации движения, не потерять окрестность пользователя)

Подход факторизации: матрица знает, что вам подойдет



Матрица факторов
пользователей:
вкусы пользователей













Матрица факторов
Предметов

NetFlix прогноз
рейтинга

Подход ранжирования (Learning to rank)

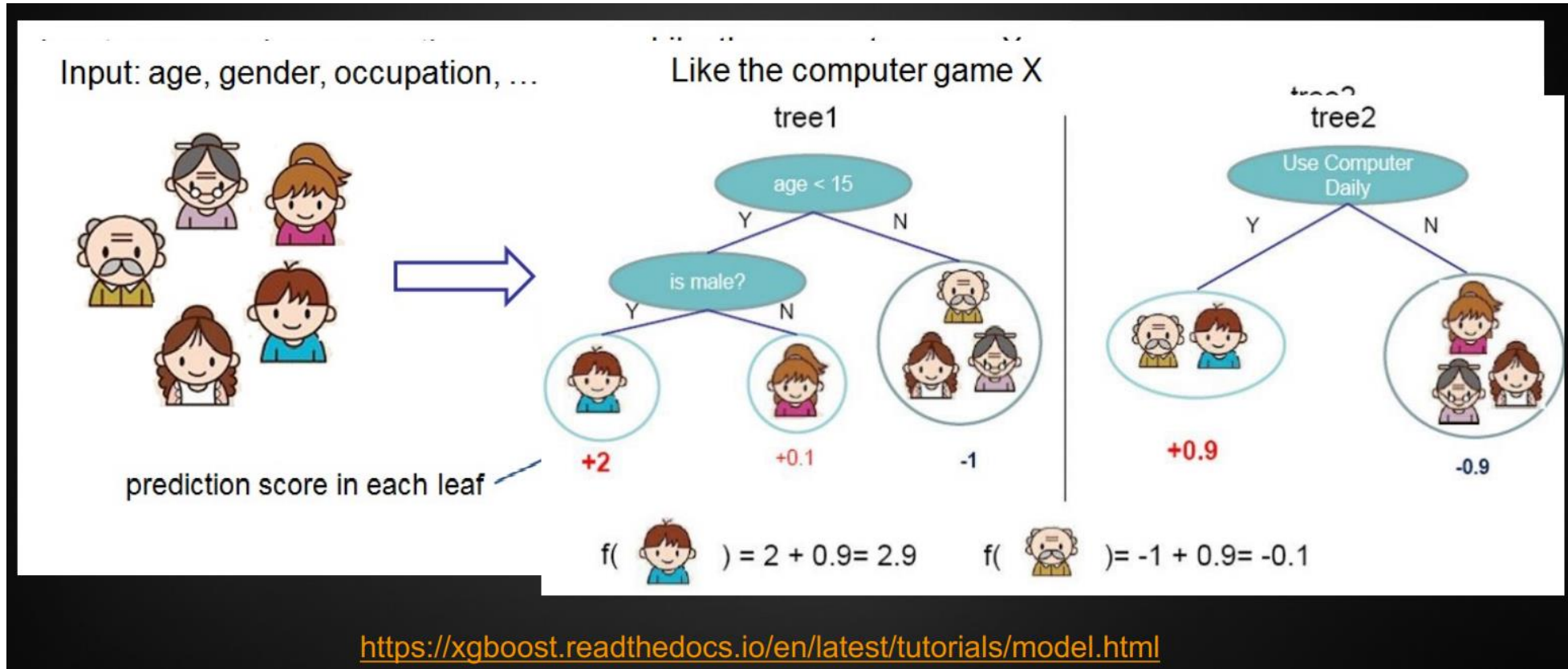
- Отсортировать список предметов, на первом месте самые интересные предметы для пользователя.
- Внедрение машинного обучения

Факторизационные машины

Пользователь			История				Тест			Цель	
											
1			1					1			1
1				1			1				1
	1		1	1					1		1
	1			1	1		1				1
	1		1		1			1			1
		1			1	1		1			1
		1		1	1					1	1
		1		1		1			1		1

- Хорошо работают с категориальными признаками

Бустинг на деревьях



- Хорошо работают с числовыми признаками, но плохо с категориальными
- Громоздки

Двухфазные рекомендательные системы



Источники информации для рекомендательных систем

Колаборативная фильтрация

- Поиск паттернов поведения в больших логах

Контентные рекомендации

- Построение профиля «интересов» пользователя

Социальные рекомендации

- Поиск рекомендаций в окружении человека

Контекстные рекомендации

- Определять что важно здесь и сейчас

Оценка эффективности рекомендательных систем

Ретроспективно, офлайн.

- есть история взаимодействия пользователя с продуктами за месяц. Для теста из этой истории берется последняя неделя, а первые три используются в качестве обучающей выборки.

Онлайн с помощью АБ-тестов.

- Оценка строится исходя из задач бизнеса: оценить разнообразие

Проблемы при подключении рекомендательных систем

Проблема холодного старта

- Новый пользователь без истории

Проблема «длинного хвоста»

- Пользователь взаимодействует только с популярными продуктами

Проблемы систем ранжирования



Feedback loop

Работа рексистемы
вносит смещение в
данные



Rich get richer

Система ротирует
контент, у которого
много позитивных
откликов.
У неизвестного, но
хорошего контента
мало шансов



Missing not at random

Пробелы в данных
могут говорить как об
отсутствии
информации, так и
быть негативным
сигналом



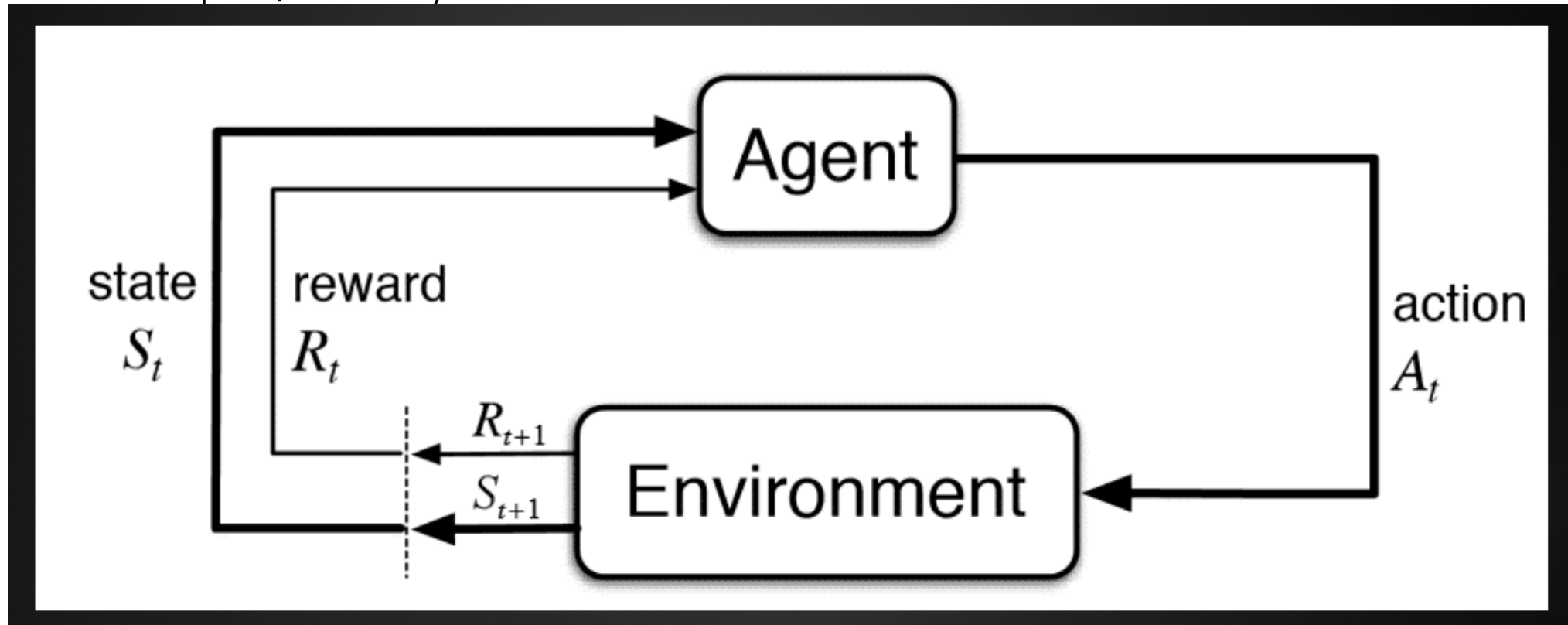
Offline-online decorrelation

С определенного
момента улучшение в
оффлайн
экспериментах не
приносят улучшения
бизнес показателей

- Рекомендательная система становится активным участником формирования исходных данных для МО
- Данные собраны в результате взаимодействия действующей сущности и среды

Обучение с подкреплением

- Обучение для действующих сущностей
- Агент общается со внешней средой и получает балы за корректное влияние на процесс обучение



Обучение с подкреплением в рекомендательных системах



- Многорукие бандиты: агент определяется в какие «игровые автоматы» кидать монетки. Запоминает какие автоматы давали награду
- Каждый автомат – это предмет (фильм), который может быть порекомендован; награда – пользователь использует или нет этот предмет (посмотрит фильм)

Контекстные многорукие бандиты

- Выбор тех предметов, от которых точно будет получена награда или тех элементов, которые мало известны
- Активно исследуется пространство действий для оптимизации вознаграждения

Примеры рекомендательных систем

Соц
сети

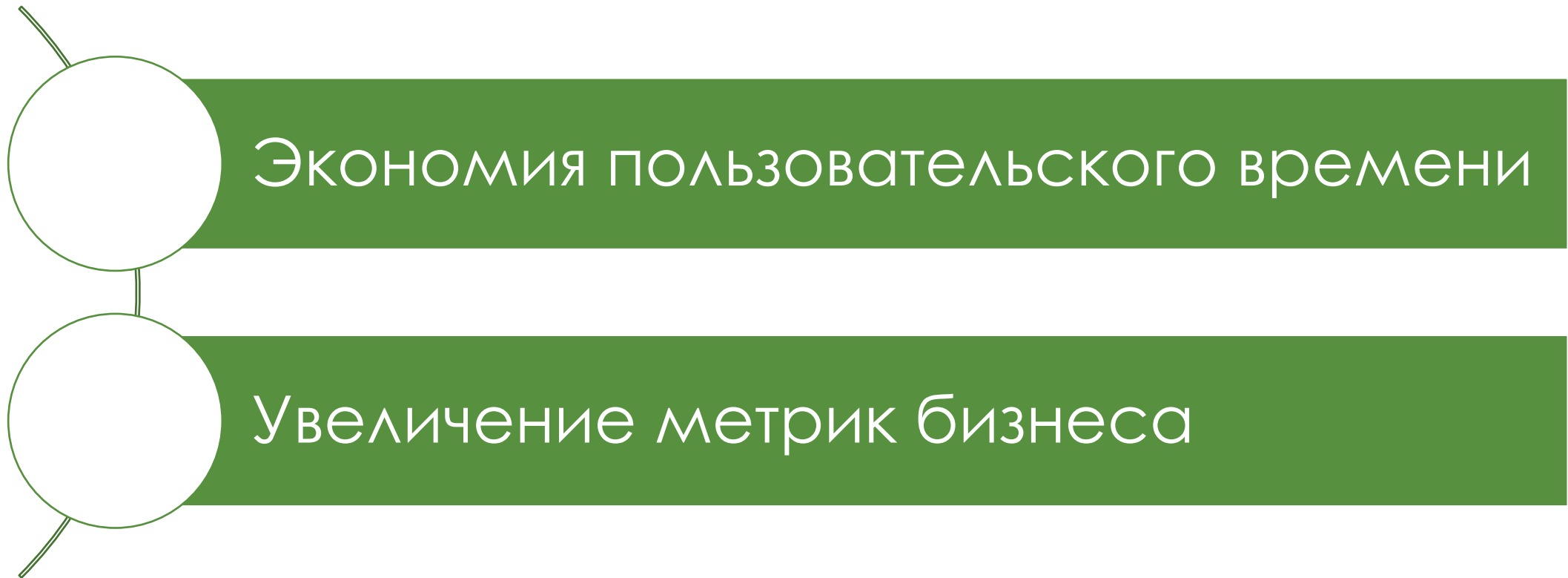
Youtube

Яндекс

Amazon

Сбер

Преимущества рекомендательных систем



Недостатки рекомендательных систем

- Она обучает пользователей не задумываться о своих желаниях, а полагаться на алгоритм. Это, в свою очередь, приводит к деградации системы, потому что ей не на чем больше обучаться.