

Виды и классификация рекомендательных систем

Цель темы

Разобрать основные виды рекомендательных систем, ограничения и проблемы, которые у них есть, и что предпочитают использовать в реальном бизнесе.

Подходы к рекомендательным системам

- ✓ Content-based-подход
- ✓ Коллаборативный подход
- ✓ Смешанный подход

Content-based-подход

Признаки у вас есть

Вы что-то знаете про User, что-то знаете про Item и что-то об их истории взаимодействия.

Решаете задачи

- Классификации
- Регрессии
- Ранжирования

Content-based-подход

Проблемы с производительностью, так как придётся скорить много айтемов.

Кроме того, не используются как таковые рекомендации, основанные на взаимодействии других юзеров.

Коллаборативные подходы

Вы будете стараться выжать всё
из взаимодействий.

Признаки бывают: **explicit** (явная оценка)
и **implicit** (просто факт взаимодействия).

User-item-матрица

	 Harry Potter	 The Triplets of Belleville	 Shrek	 The Dark Knight Rises	 Memento
	✓		✓	✓	
		✓			✓
	✓	✓	✓		
				✓	✓

Коллаборативные подходы

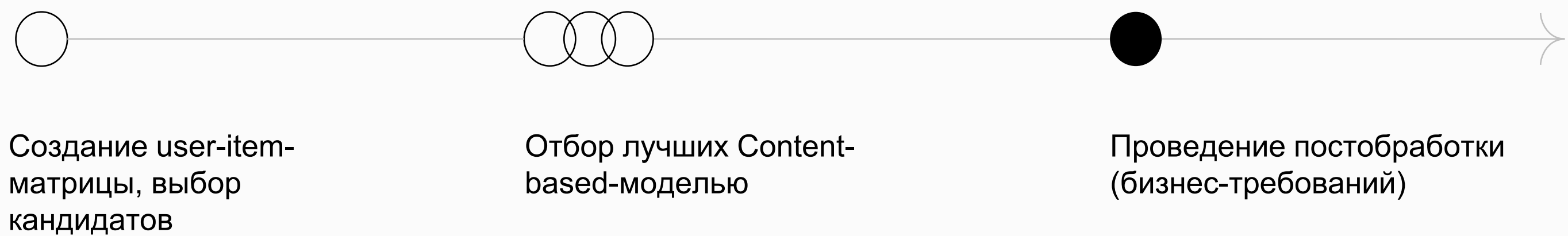
Существенный минус заключается в том, что вы не используете информацию о ваших юзерах и айтемах для тонкой подстройки, но этот подход всё же хорошо подходит для первоначального отбора.

Двухэтапные рекомендации

Почти во всех реальных системах используются так называемые двухэтапные рекомендации. Также такие системы называют **гибридными**.

Кандидаты — первоначальный список, выбранный рекомендательной системой на первом этапе.

Последовательность действий



Двухэтапные рекомендации

Необходим препроцессинг для сокращения количества данных и постпроцессинг для учёта требований бизнеса.

Можно отфильтровать:

1. Айтемы, с которыми мало взаимодействовали
2. Юзеров, у которых было мало взаимодействий с айтемами

Холодный старт

Бывает, что у вас почти нет информации о том, как пользователь взаимодействовал с айтемами на сервисе.

Особенно если это новый юзер, поэтому возникает проблема «холодного старта».

Холодный старт

Для пользователя

Что делать, если к вам приходит новый пользователь?

- Самое популярное
- Самое рейтинговое
- А, может, спросить?

Холодный старт

Для айтема

А что, если новый айтем?

- Случайный показ
- Показ на основе близости

Длинный хвост

Проблема «длинного хвоста» пересекается с проблемой холодного старта для айтемов: товаров слишком много — как правильно показывать?

Статистическая значимость

Проблема статистической значимости
достаточно общая для всех задач ML.

Вывод темы

Сегодня вы познакомились с тем, какие рекомендательные системы используются в реальных сервисах и что такое проблема «холодного старта».