

Гибридные архитектуры

Введение



В этом уроке мы поговорим о том как глобально сочетать разные подходы.

План

1

Общее понятие
гибридной
архитектуры

2

Content-based +
Collaborative
filtering

3

Разные типы
гибридных
архитектур

Общее понятие гибридной архитектуры

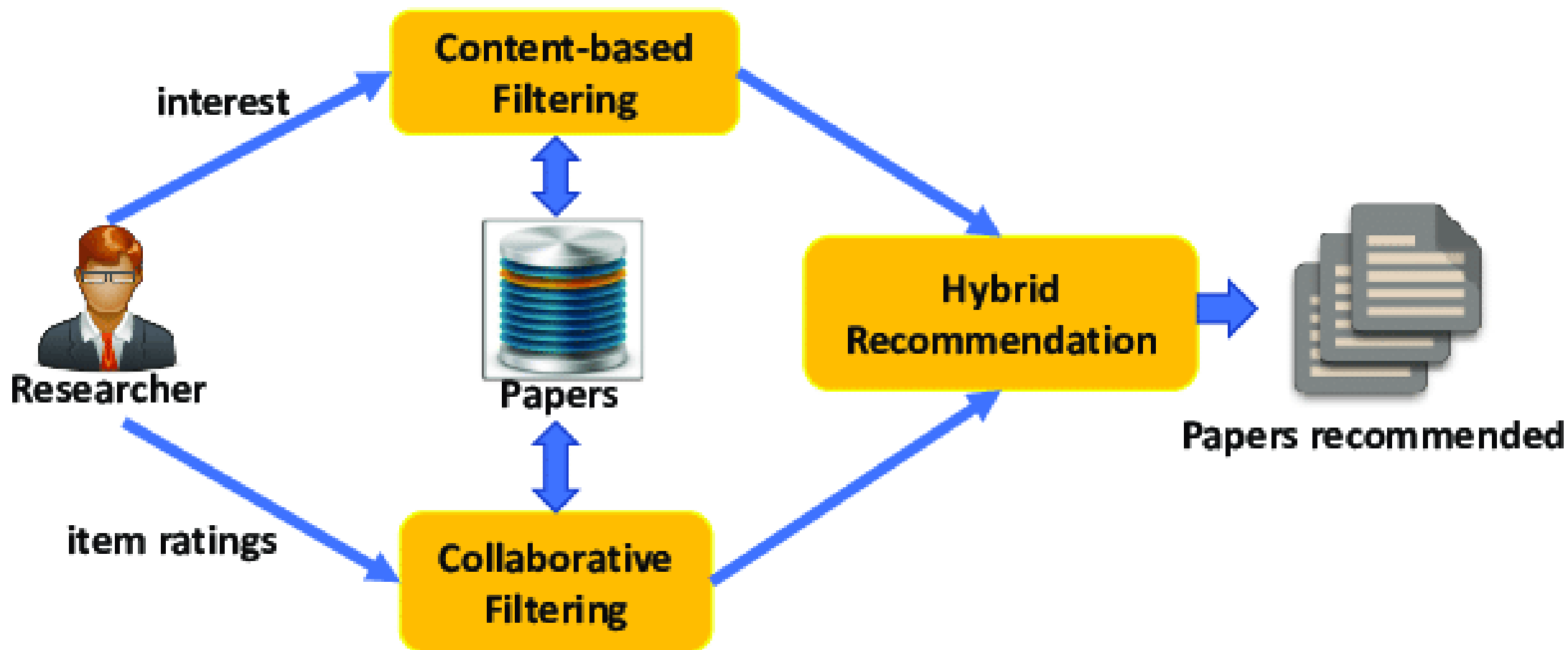
Гибридный подход

Гибридный подход – это сочетание коллаборативной и контентной фильтраций.

- Улучшает качество рекомендаций
- Позволяет сначала взвешивать результаты согласно контентной фильтрации, а затем смещать эти веса по направлению к коллаборативной фильтрации (по мере "вызревания" доступного набора данных по конкретному пользователю).



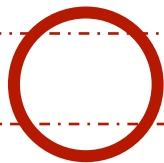
Гибридный подход



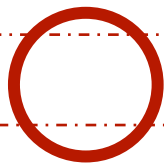
Content-based + Collaborative filtering

**вернемся к 1-ому
уроку...**

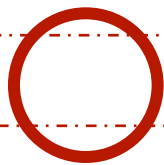
Виды рекомендаций



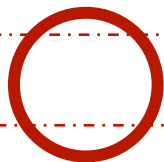
Коллаборативная фильтрация (collaborative filtering)



Основанные на контенте (content-based)



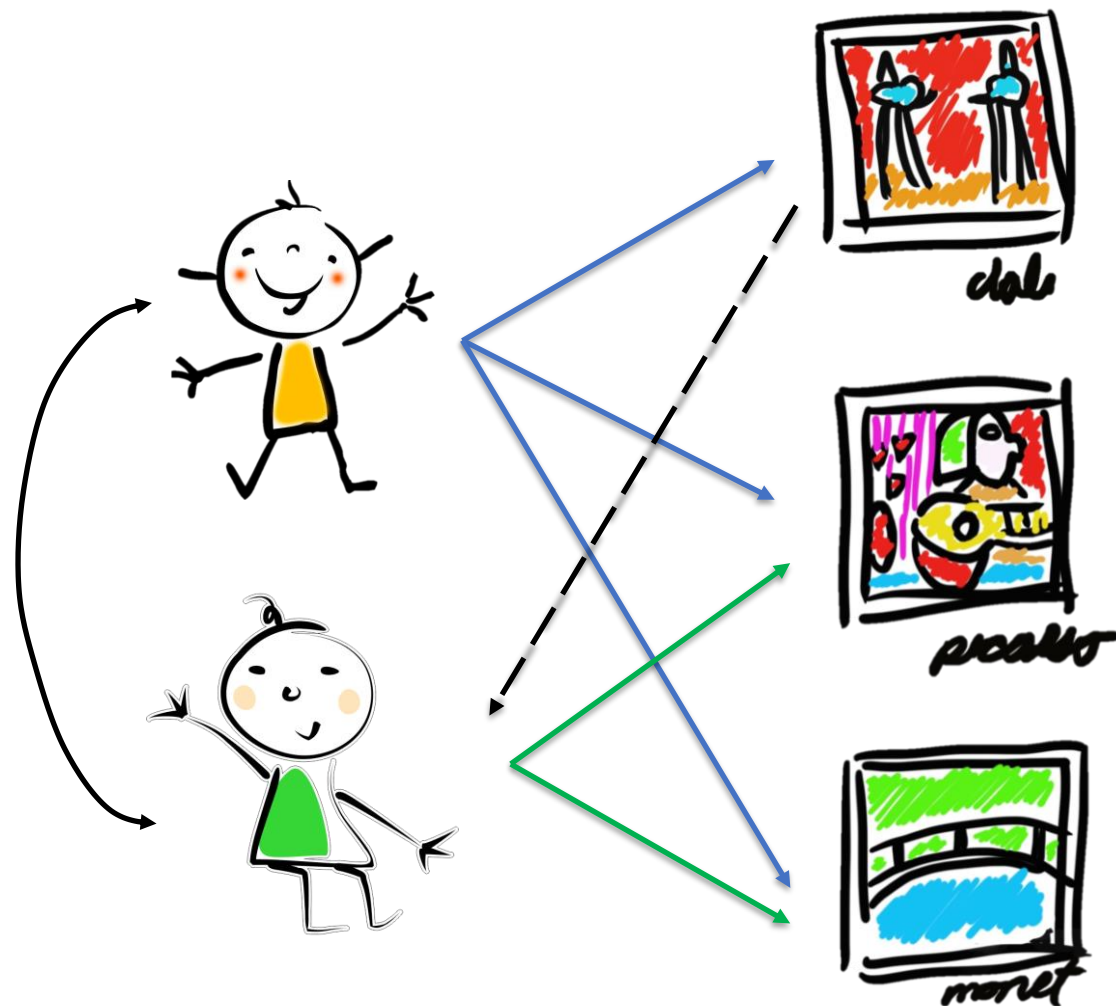
Основанные на знаниях (knowledge-based)



Гибридные (hybrid)

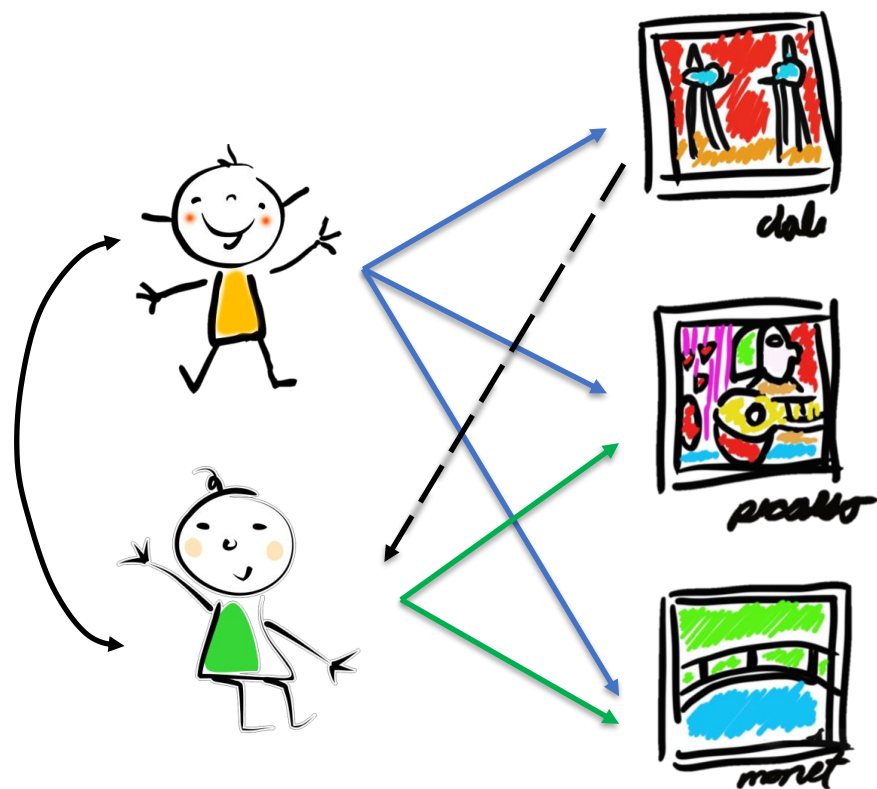
Коллаборативная фильтрация

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

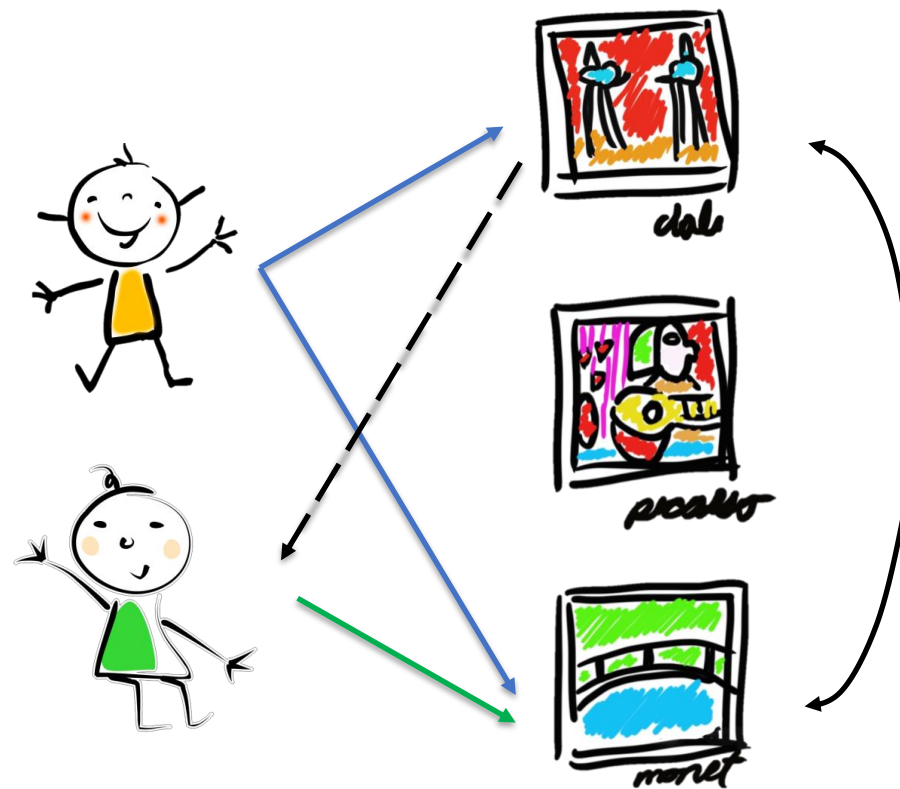


Коллаборативная фильтрация

1 User-based

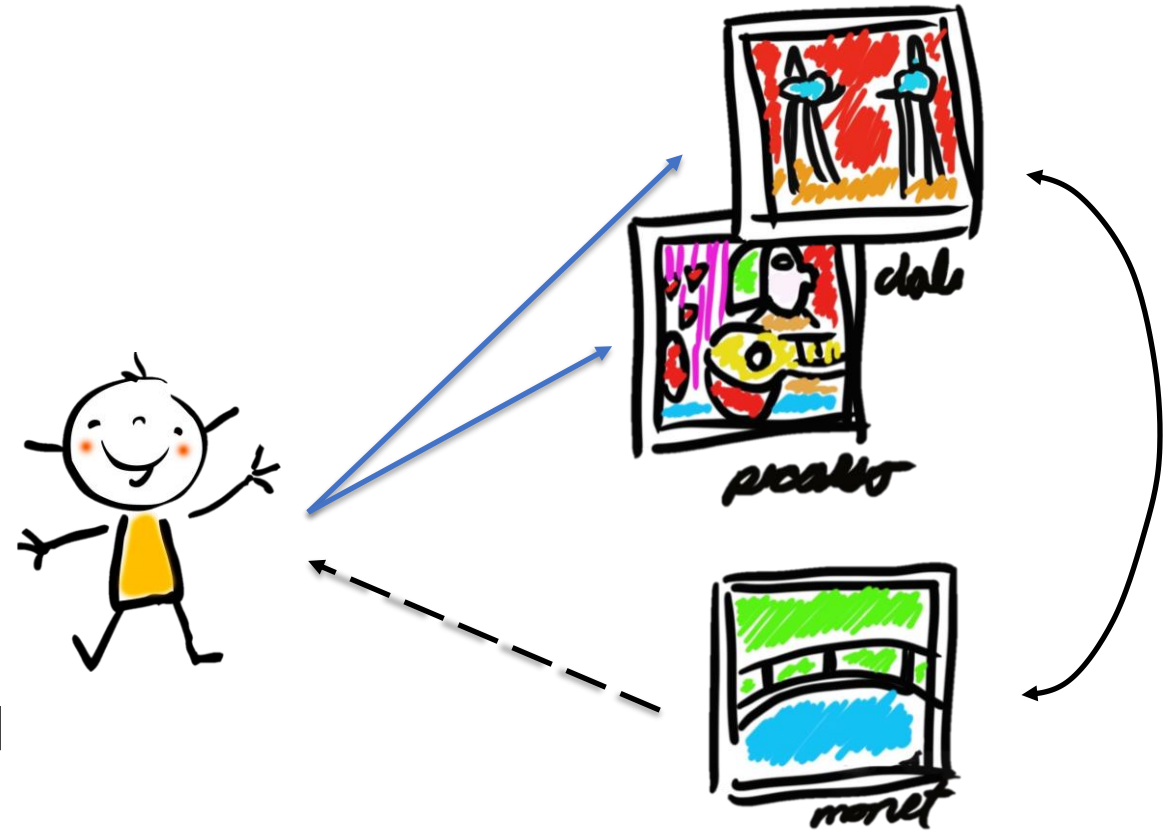


2 Item-based



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.

Пример с сайта reStore

Созданы друг для друга



Apple iPhone 11 Pro Max, 256 ГБ, серебристый
106990 Р

+ **- 500 Р**

Чехол Apple для iPhone 11 Pro Max, силикон, (PRODUCT)RED
2890 Р 3390 Р

= **109880 Р** 110380 Р

Купить вместе

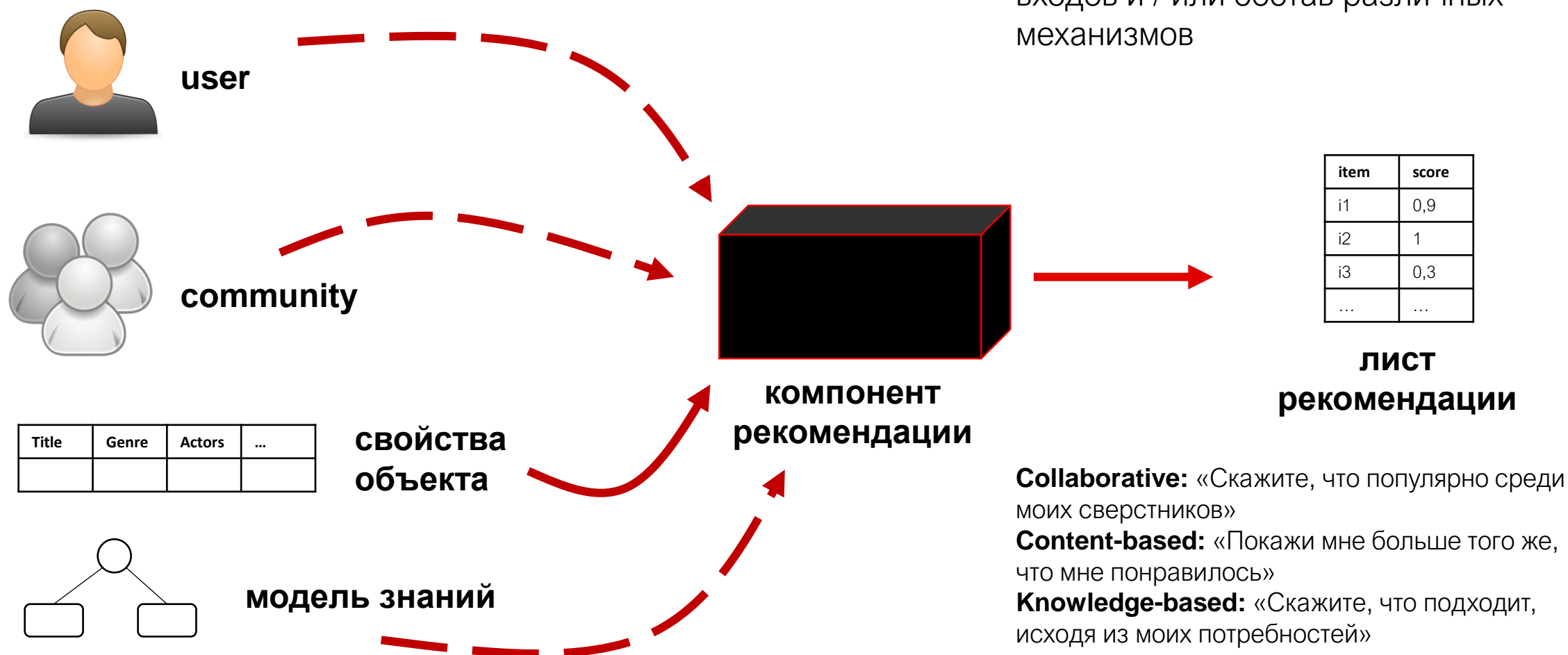
Гибридные



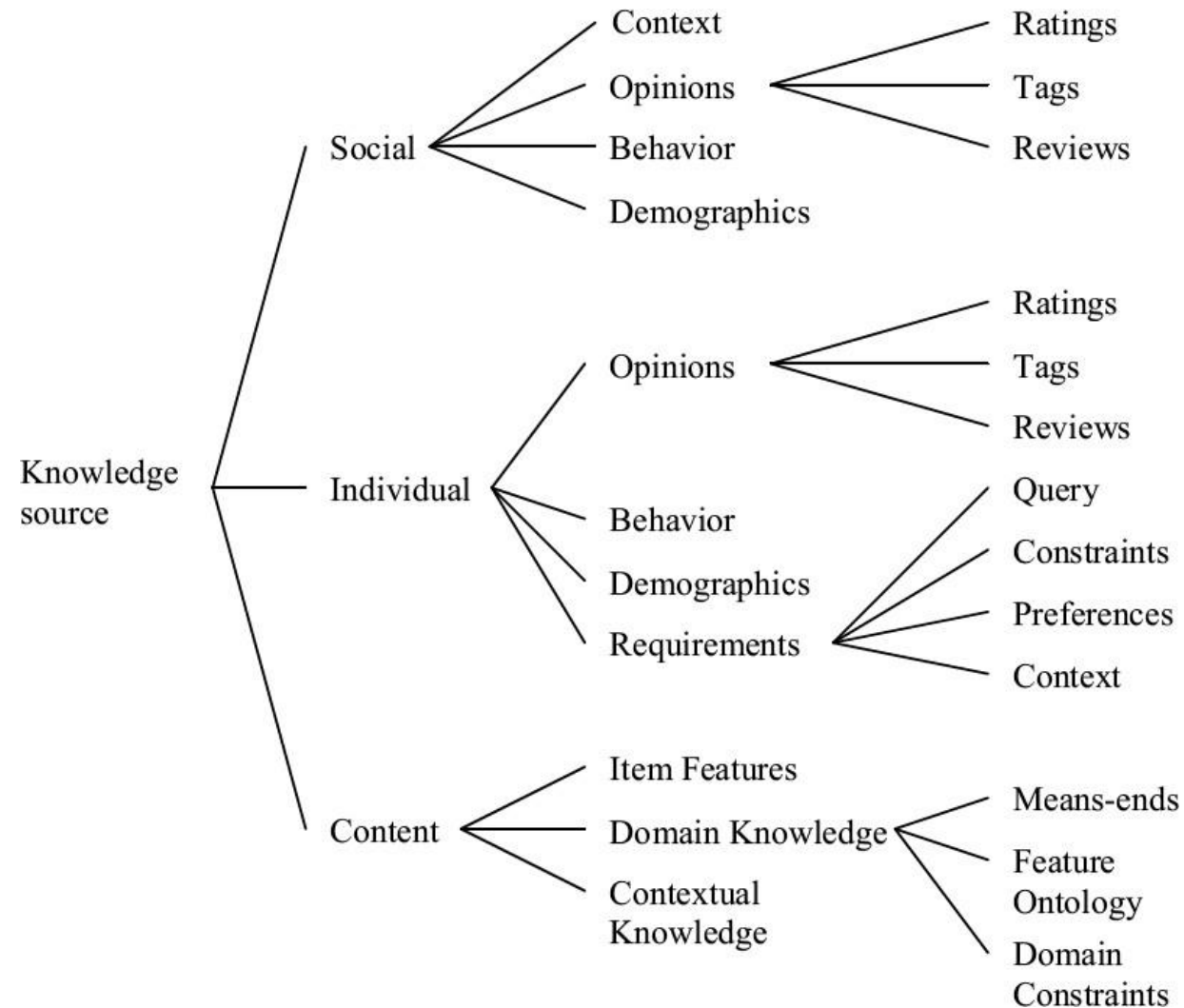
Распространённые типы комбинирования:

- 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

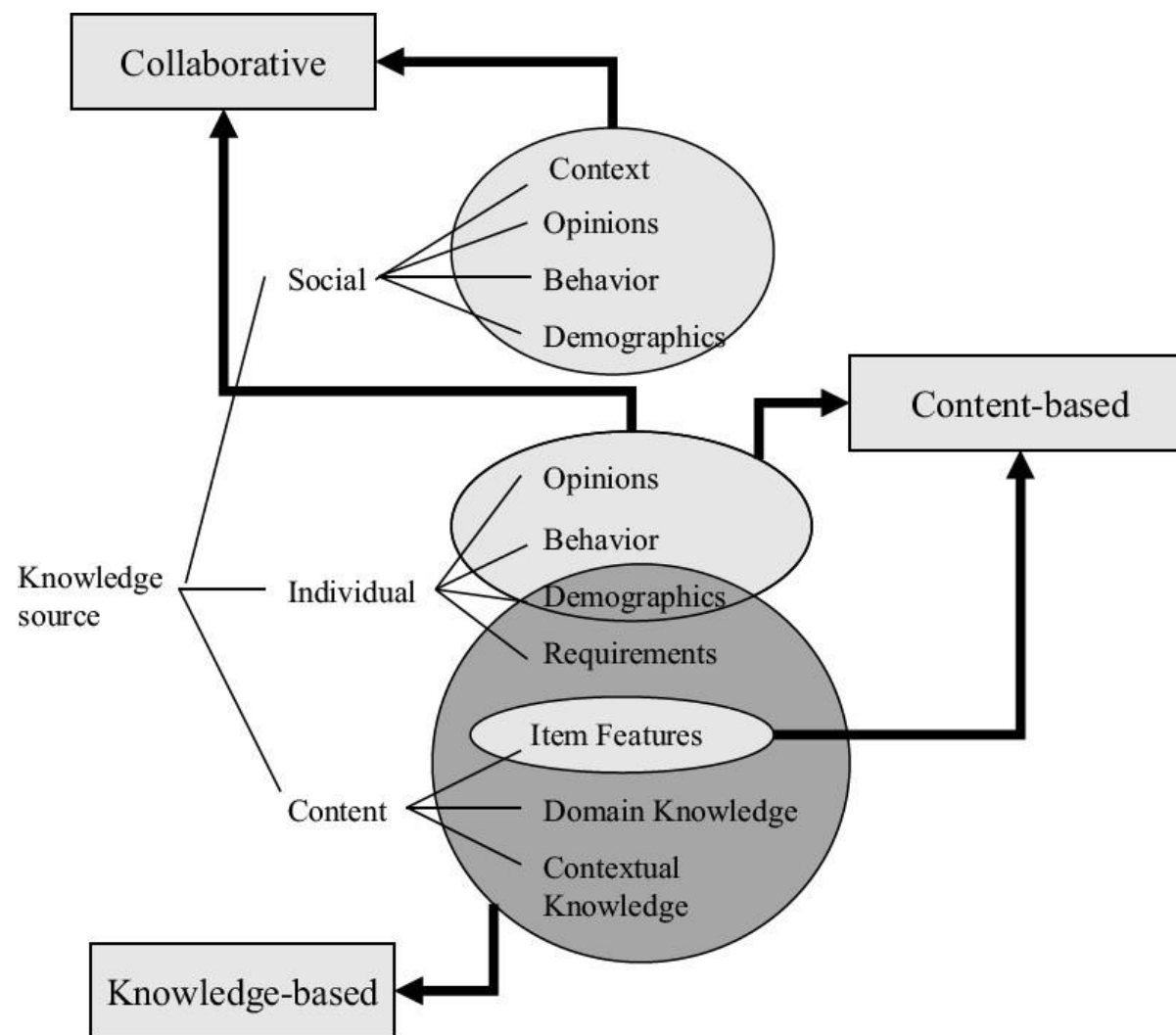
Гибридный подход



Таксономия источников знаний

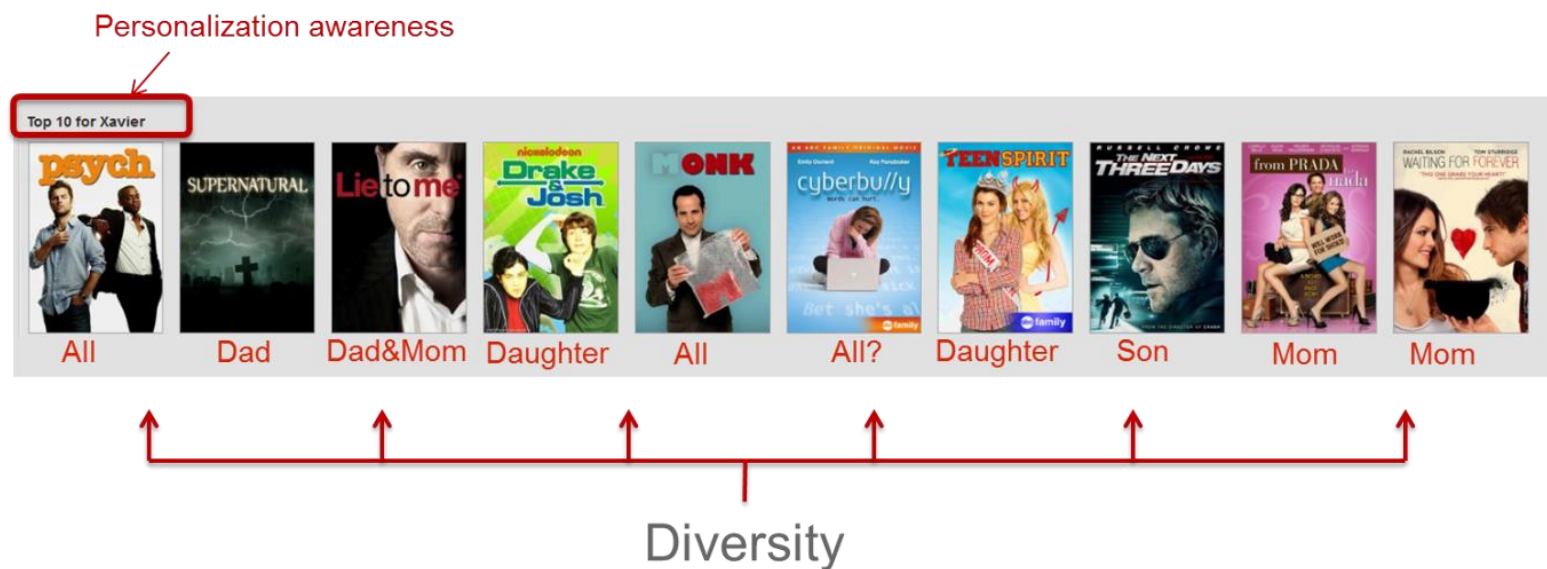


Источники знаний и типы рекомендаций



Зачем строят гибриды?

- Разные подходы имеют свои недостатки и преимущества
- Вместо выбора – гибко используем все



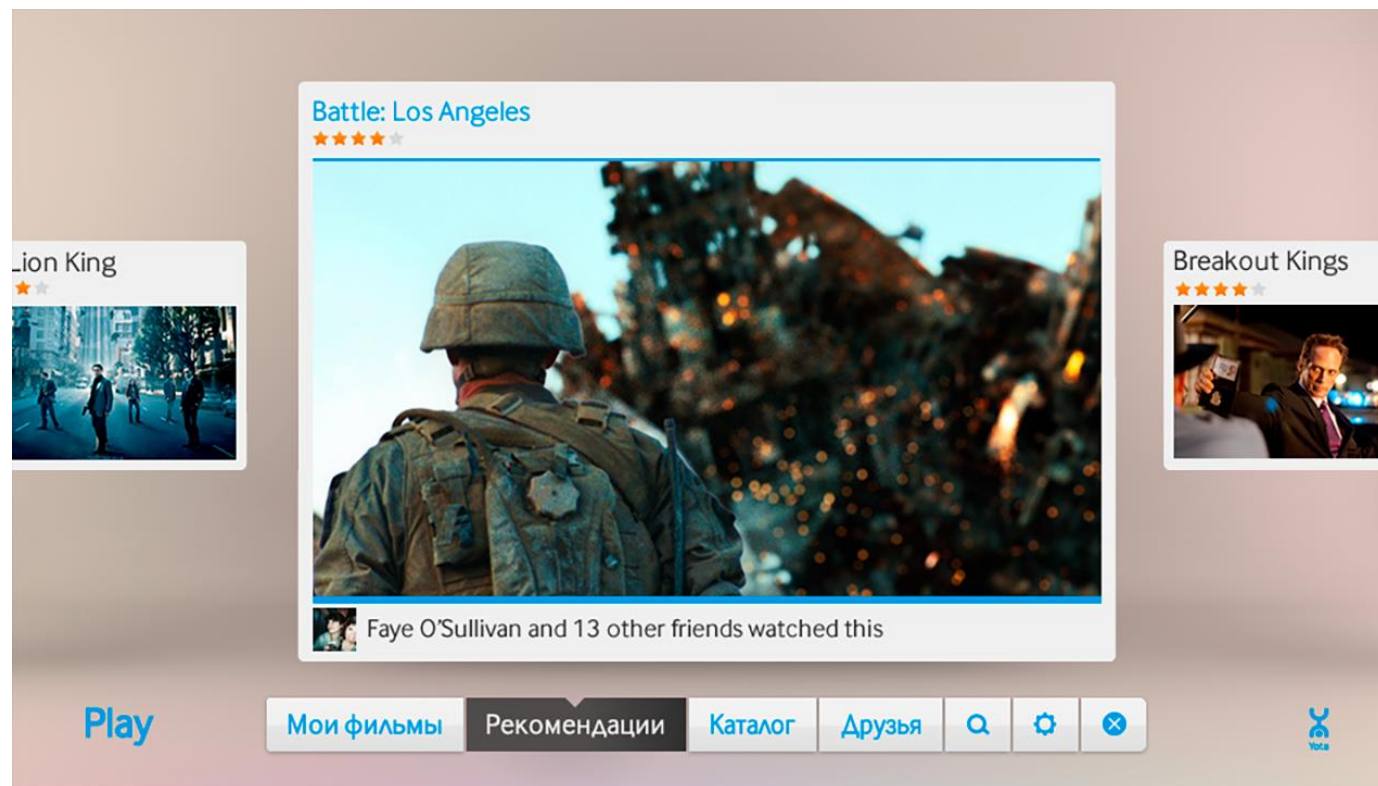
Гибридизация



- Часто улучшает качество рекомендаций
- Иногда положительно сказывается на разнообразии
- Не гарантирует решение всех проблем, связанных с тем или иным подходом

Пример – рекомендации в Okko

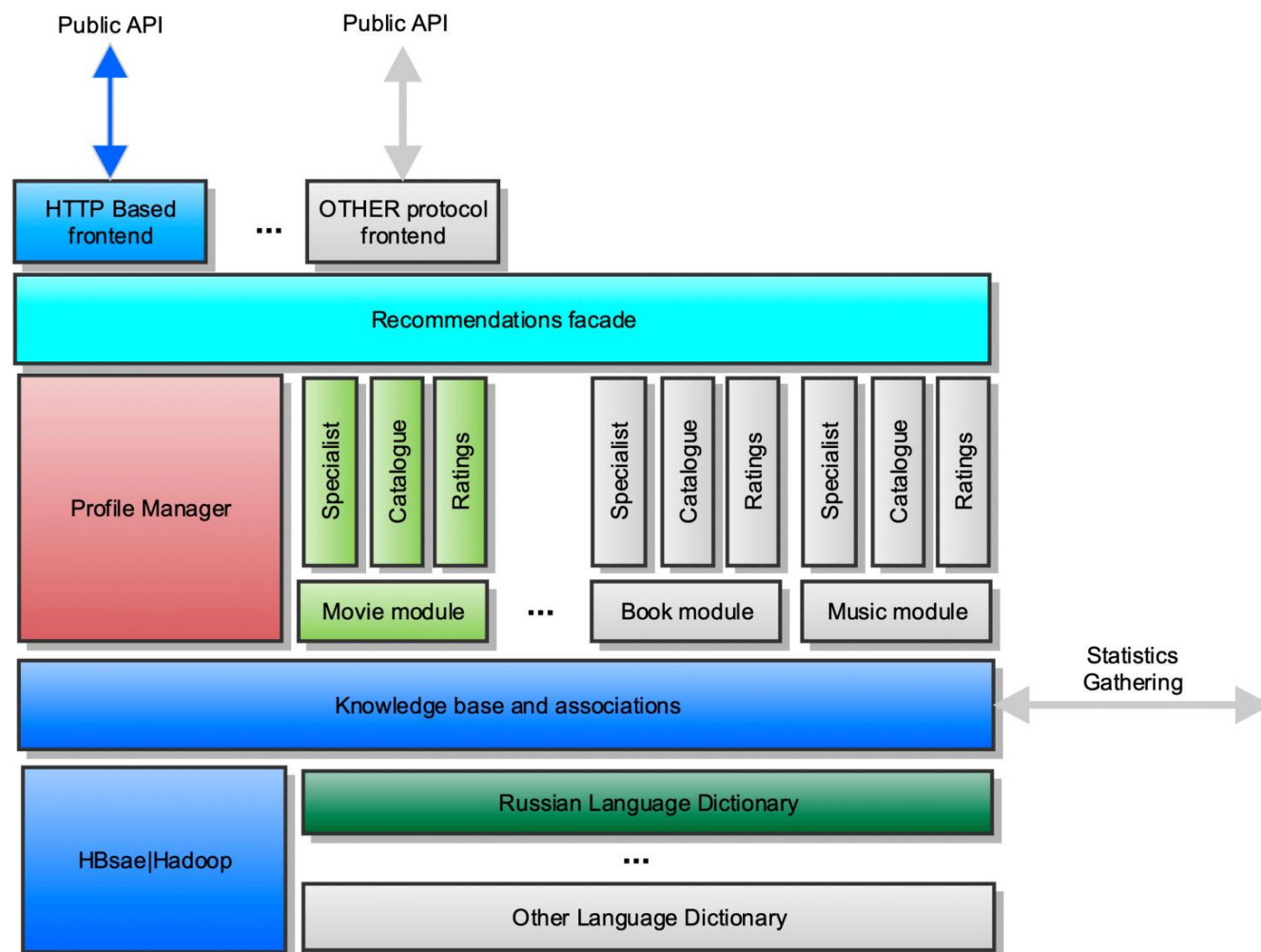
Okko начал своё существование в 2011 году как часть Йоты, запустившись под именем Yota Play



<https://habr.com/ru/company/okko/blog/454224/>

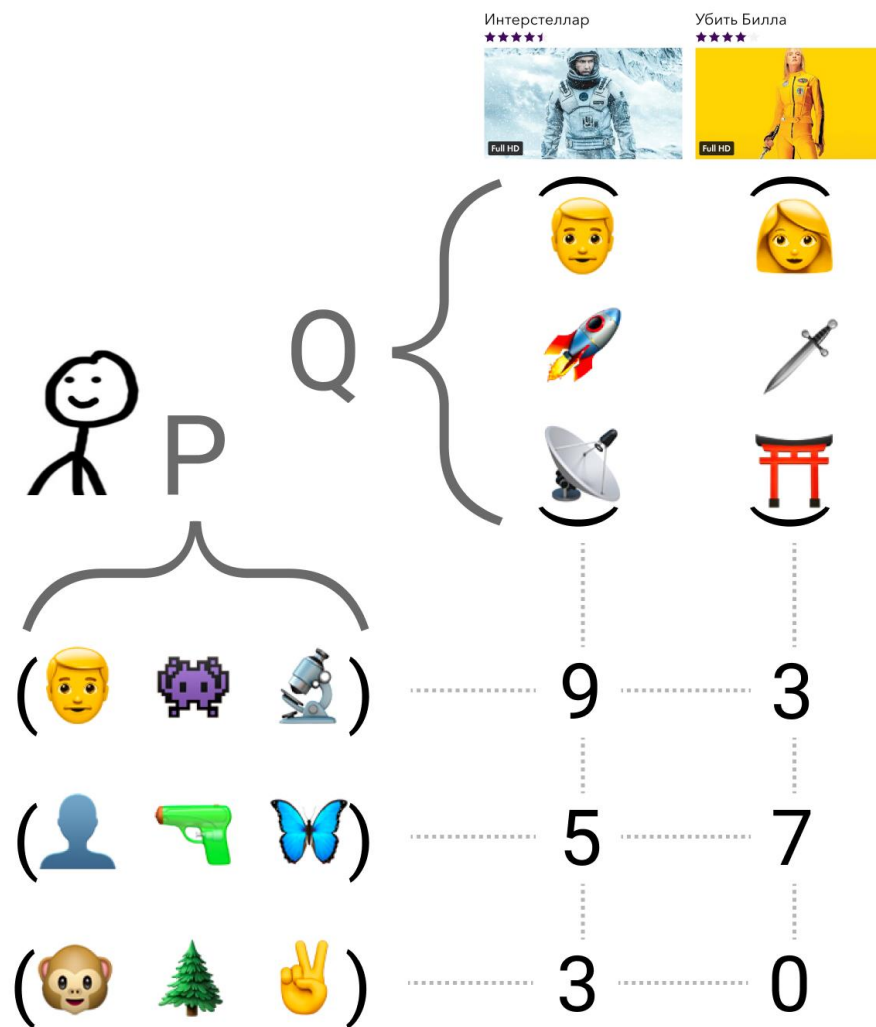
Рекомендации в Okko

В 2012 социальные рекомендации было решено дополнить алгоритмическими. Так появился **«Оракул»** — первая рекомендательная система онлайн-кинотеатра Okko.



Рекомендации в Okko

Решение: Комбинация матричной факторизации с WARP loss и градиентного бустинга над деревьями



Weighted Approximate-Rank Pairwise

Идея: Использовать adaptive learning rate (поговорим об этом в лекции об active learning). Будем оценивать качество, исходя из количества семплов, которые нам пришлось бы просмотреть, чтобы для данной пары {пользователь, положительный пример} найти отрицательный пример, который система оценила выше положительного.

Чем раньше мы находим такой пример – тем хуже.

$$L_{WARP}(u, i, j) = \frac{\hat{L}(\text{rank}_u^1(i))}{\hat{L}(|Items|)} \cdot (\hat{r}_{uj} + 1 - \hat{r}_{ui})$$

$$\hat{L}(k) = \sum_{l=1}^k \frac{1}{l}$$

$$\text{rank}_u^1 \approx \frac{|Items| - 1}{\text{numdraws}(j)}$$

<https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/37180.pdf>

<https://www.ee.columbia.edu/~ronw/pubs/recsys2013-kaos.pdf>

REKKO CHALLENGE 2019



Соревнование от OKKO: https://boosters.pro/championship/rekko_challenge/data

Задача

На предоставленном тренировочном множестве обучить рекомендательную систему и предсказать топ 20 наиболее релевантных для пользователя идентификаторов контента. Обучающее множество собрано за N дней ($N > 60$), тестовое множество — за последующие 60 дней.

Целевая переменная

В качестве целевой переменной необходимо предсказать множество фильмов, которые пользователь потребил за тестовый период.

<https://habr.com/ru/company/tinkoff/blog/454818/>

<https://www.youtube.com/watch?v=-eCr1K9IKxg>

Разные типы гибридных архитектур

Виды гибридизации



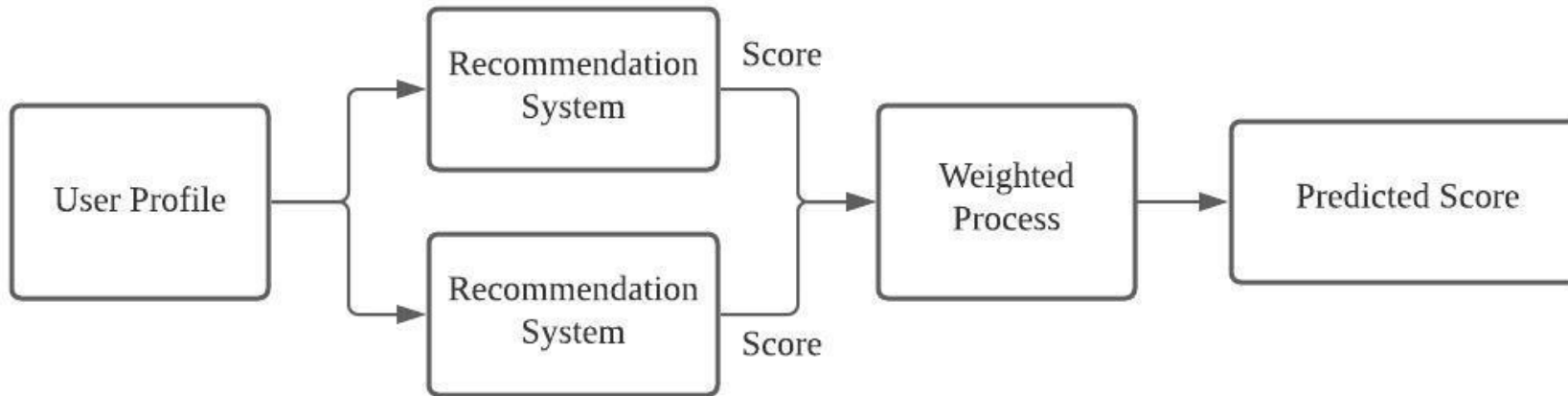
- Weighted
- Switching
- Mixed
- Feature augmentation
- Cascade
- Feature Combination

Weighted



- Рекомендации строятся на основе комбинирования оценок от разных систем с весами
- Стратегия объединения: считать средневзвешенный прогноз по нескольким оценкам
- Например:
 - Линейная комбинация
 - Голосование

Weighted

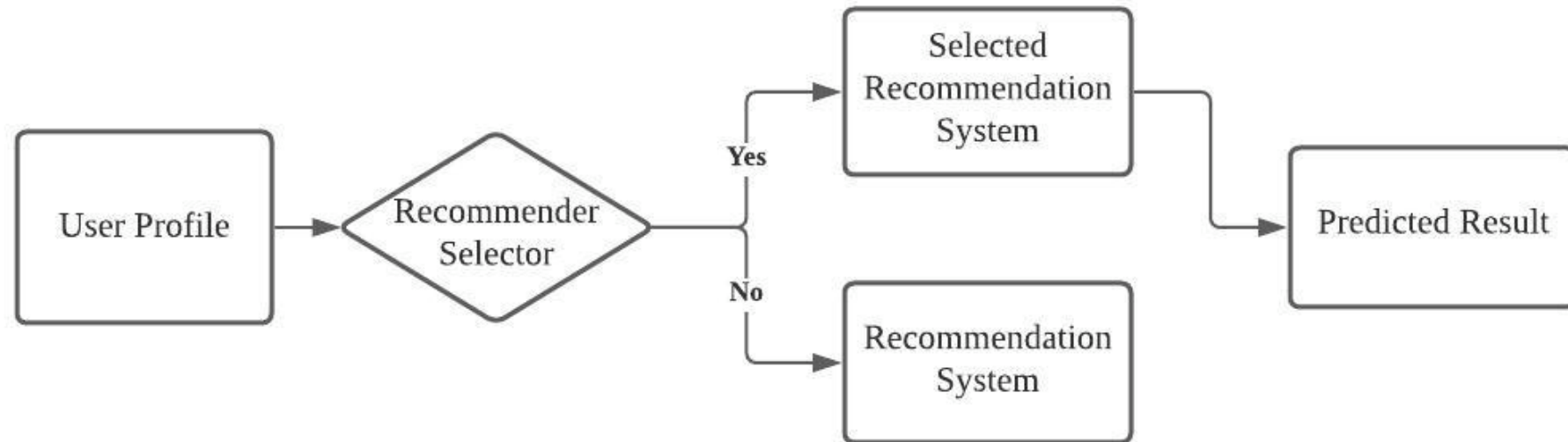


Switching



- Рекомендации строятся путем переключения между системами, работающими независимо, на основании критериев для переключения
- Стратегия объединения: для разных продуктов/пользователей применять различные алгоритмы,

Switching

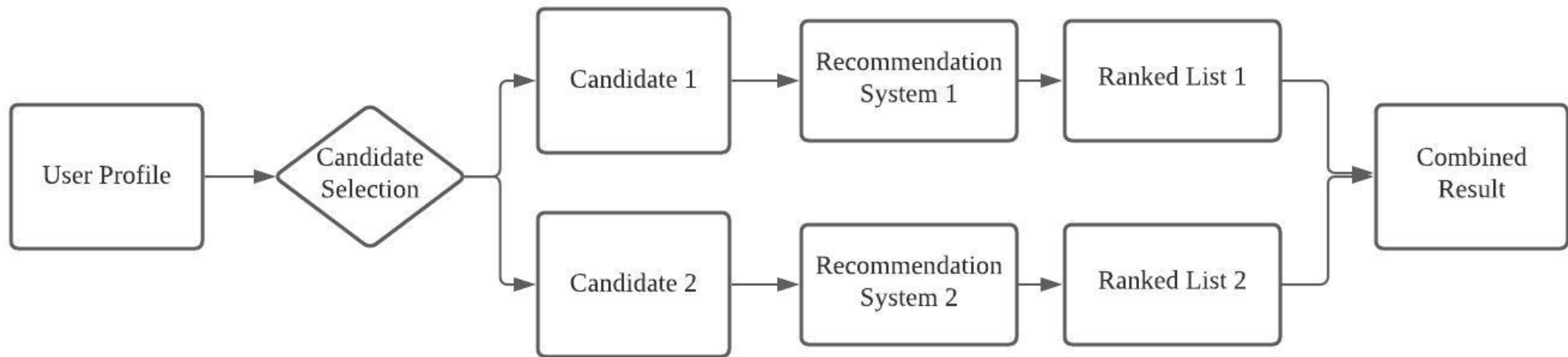


Mixed



- Список рекомендаций состоит из «смеси» рекомендаций от разных систем
- Стратегия объединения: вычисляются рекомендации по разным алгоритмам, а потом просто объединяются в один список.

Mixed

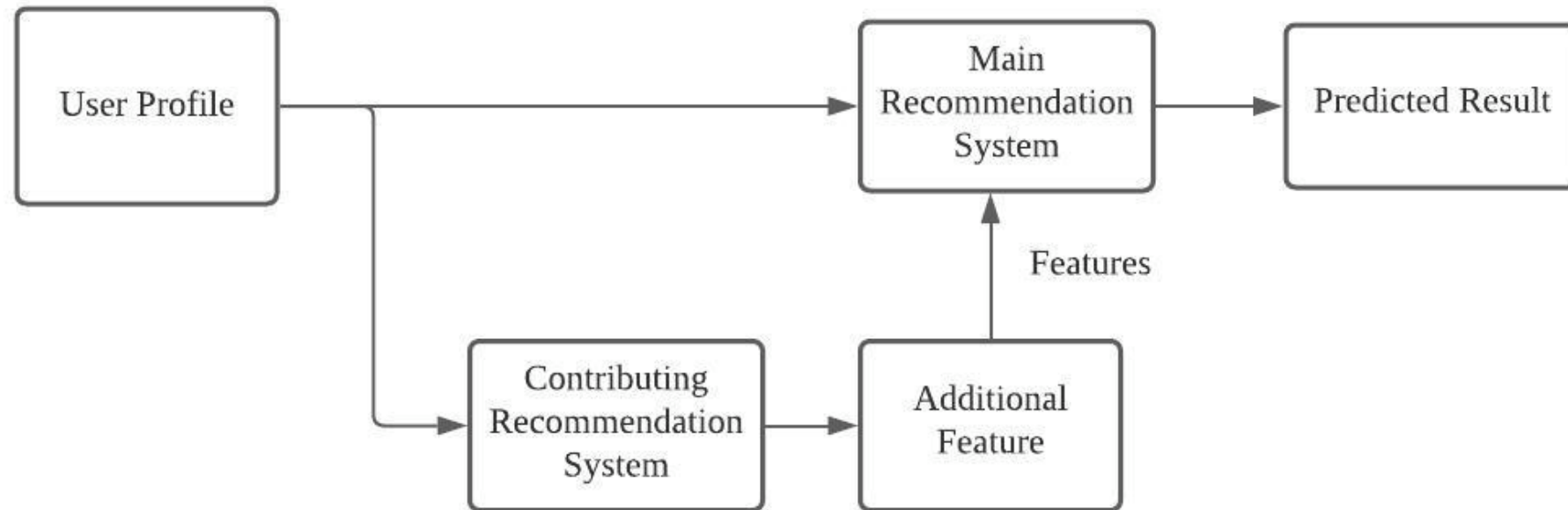


Feature Combination



- Подход основан на content-based
- Признаки от разных систем объединяются в одну выборку для построения единой модели

Feature Combination



Feature Augmentation



- Модель рекомендаций используется для создания рейтинга или классификации профиля пользователя / элемента, которая в дальнейшем используется в основной системе рекомендаций для получения окончательного прогнозируемого результата.

Feature Augmentation

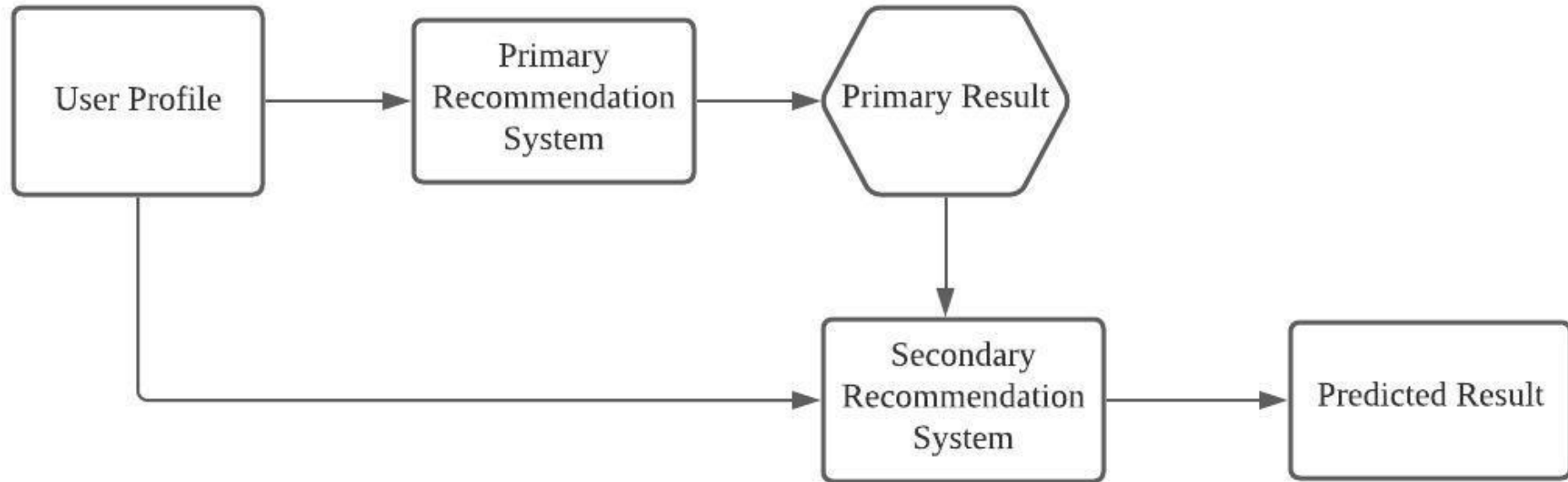


Cascade



- Поэтапное применение нескольких моделей для уточнений рекомендаций
- Candidate selection

Cascade



Вопросы

Заключение



- Гибриды не хуже чем «стандартные» подходы
- Есть разные виды гибридов

Семинар:

1. Разбор домашних работ №1
2. Как совместить CF + CB?