

СРАВНИТЕЛЬНЫЙ АНАЛИЗ АЛГОРИТМОВ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Рекомендательные системы – это программы или веб-сервисы, целью которой является попытка предсказать, какие объекты будут интересны пользователю, основываясь на данных его профиля.

Коллаборативная фильтрация – это технология прогнозирования предпочтений пользователя с учётом интересов других посетителей ресурса. На основе собираемой информации система рекомендует те товары, которыми аудитория со схожими интересами уже интересовалась, а конкретный человек – ещё нет.

В основе метода лежит использование данных, полученных от пользователей.

Алгоритм содержит следующие шаги:

- построение матрицы оценок;
- выбор метрики сходства (обычно косинусная мера);
- оценка сходства для выбранных пользователей;
- распределение весов;
- расчёт суммы калиброванных оценок для каждого объекта.

Преимущества коллаборативной фильтрации:

- не требует знаний о предметной области и работы человека;
- возможны предложения новых объектов, ранее не знакомых конкретному пользователю;
- персонализированные рекомендации.

Недостатки коллаборативной фильтрации:

- разрежённость матрицы оценок;
- сложность расчёта для большого количества объектов;
- холодный старт;
- синонимия.

Фильтрация на основе содержания – это метод рекомендательных систем, который использует информацию о характеристиках объектов, на основе которых делает предположение о том, понравится ли пользователю данный объект.

Алгоритм фильтрации выглядит следующим образом:

- сбор данных, характеризующих объект;
- построение профиля пользователя;
- поиск сходства между элементами.

Как и в случае коллаборативной фильтрации, могут использоваться различные метрики, но чаще всего предпочтение отдаётся косинусной мере. Необходимо рассчитать разницу между векторами пользователя и объекта, чем меньше эта разница, с тем большей вероятностью пользователю понравится рассматриваемый объект.

Преимущества фильтрации на основе содержимого:

- легко масштабируется
- рекомендации основываются на повседневной деятельности пользователей, что повышает точность рекомендаций;
- все новые объекты с заданными характеристиками могут начинать рекомендоваться сразу после их появления.

Недостатки фильтрации на основе содержимого:

- ограниченность рекомендаций характеристиками;
- проблема холодного старта;
- не учитываются контекст и персональные предпочтения;
- субъективность характеристик.

Методы фильтрации, основанные на контенте, требуют значительного объёма информации о функциях элемента, а не о его взаимодействии с пользователем. Коллаборативная фильтрация, с другой стороны, использует взаимодействия между пользователями и объектами, чтобы сгруппировать пользователей со схожими вкусами и предложить целевому пользователю новые объекты, которые нравятся группе.

Модели фильтрации на основе контента в значительной степени основаны на знании предметной области, поскольку функции элементов встроены в систему вручную. Для совместной фильтрации не требуется такого глубокого знания предметной области, поскольку все встраивания выполняются автоматически.

Системы коллаборативной фильтрации требуют только данных о поведении пользователя, в то время как методы, основанные на содержании, требуют данных как о пользователе, так и об объекте.

Коллаборативная фильтрация эффективнее в случае, когда есть обширные данные о взаимодействиях пользователей с элементами, но эти данные могут быть разреженными. Фильтрация на основе содержимого, напротив, более эффективна тогда, когда проще получить данные о свойствах объектов, нежели о взаимодействии с ними.

Оба типа фильтрации могут использоваться в схожих областях, так что выбор конкретного зависит от свойств рассматриваемых объектов, от предметной области и целей.