

Коллаборативная фильтрация

Дмитрий Шипилов

ML Engineer, Uzum Market

Skillbox

Задачи модуля

- 1 Узнать, как работает KNN для рекомендательных систем
- 2 Изучить метрики рекомендательных систем
- 3 Понять, как использовать матричные разложения

В результате модуля ВЫ...

- ✓ Узнаете, как применять коллаборативную фильтрацию к задаче рекомендаций
- ✓ Изучите границы применения для метода KNN и как его можно улучшить
- ✓ Изучите метрики, которые чаще всего используются в рекомендательных системах, и узнаете, как их подбирать

Метод ближайших соседей

Недостатки content-based-подхода

- Сложность при больших выборках
- Сложность при получении признаков для некоторых областей
- Плохо учитывается опыт других пользователей

KNN-подход

KNN: k nearest neighbors — метод K ближайших соседей

- Основан на «схожести» векторов
- Можем использовать любую метрику расстояния
- Как определить, какой сосед «близкий» и сколько их должно быть

User-item-матрица

	Бесконечная Шутка	Война и Мир	Гарри Поттер и Орден Феникса	Улисс	Финансист
User1	5	0	3	5	4
User2	0	5	3	0	5
User3	0	3	5	1	0
...
UserN	5	3	1	4	4

Explicit — когда у вас есть явные оценки от пользователя для айтема.

Implicit — когда у вас есть только факты взаимодействия пользователя и айтема.

Пользователь к пользователю

	Бесконечная Шутка	Война и Мир	Гарри Поттер и Орден Феникса	Улисс	Финансист	
User1	5	0	3	5		4
User2	0	5	3	0		5
User3	0	3	5	1		0
...	
UserN	5	3	1	4		4

Item к Item

	Бесконечная Шутка	Война и Мир	Гарри Поттер и Орден Феникса	Улисс	Финансист
User1	5	0	3	5	4
User2	0	5	3	0	5
User3	0	3	5	1	0
...
UserN	5	3	1	4	4

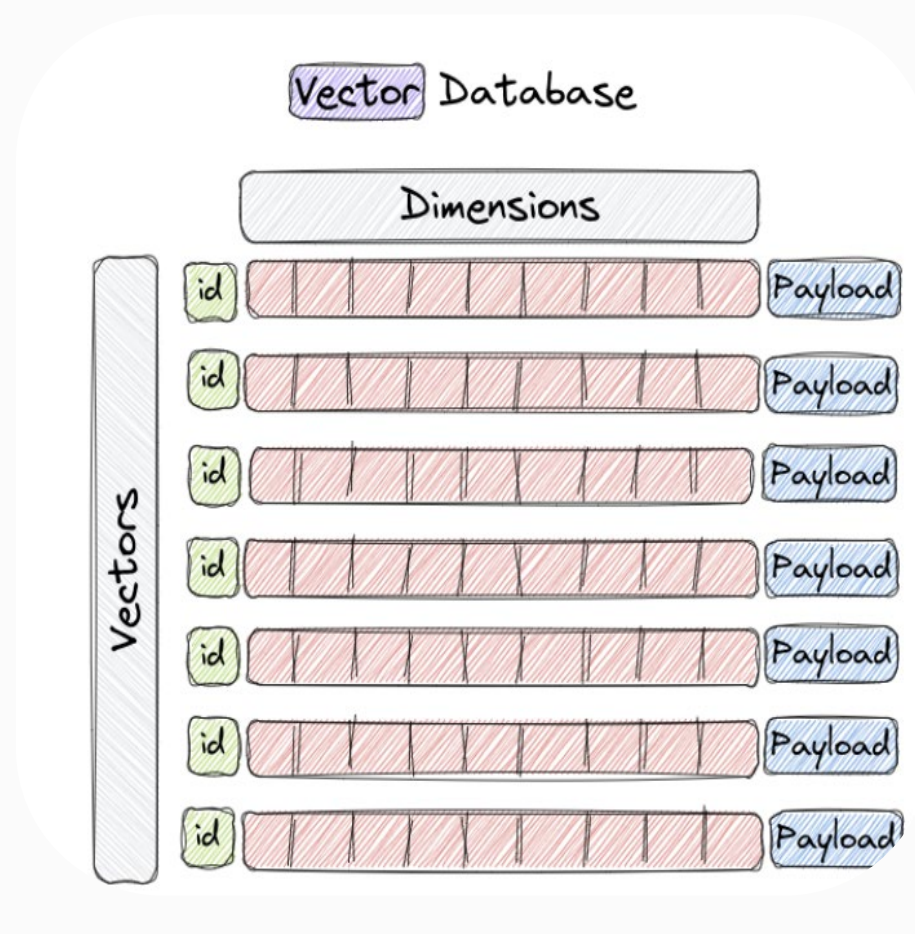
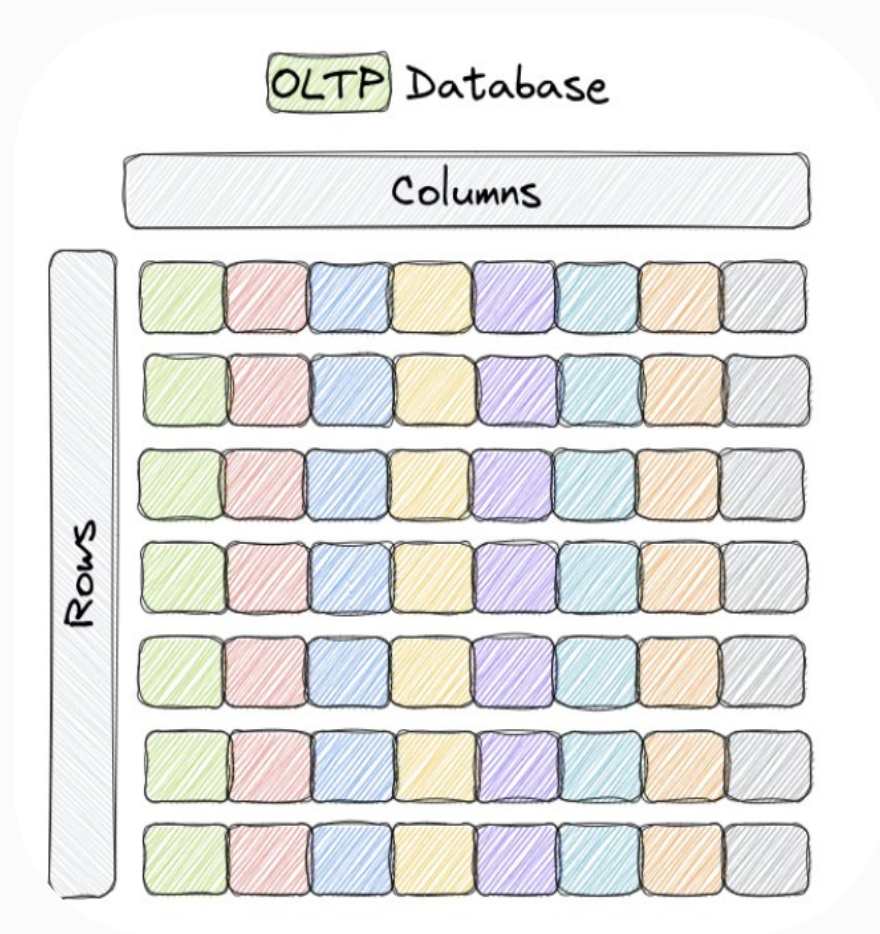
Ограничения user-item-матрицы

- Без фильтрации размер матрицы слишком большой
- Для хранения подобных матриц используется специальный инструмент Sparse из пакета SciPy

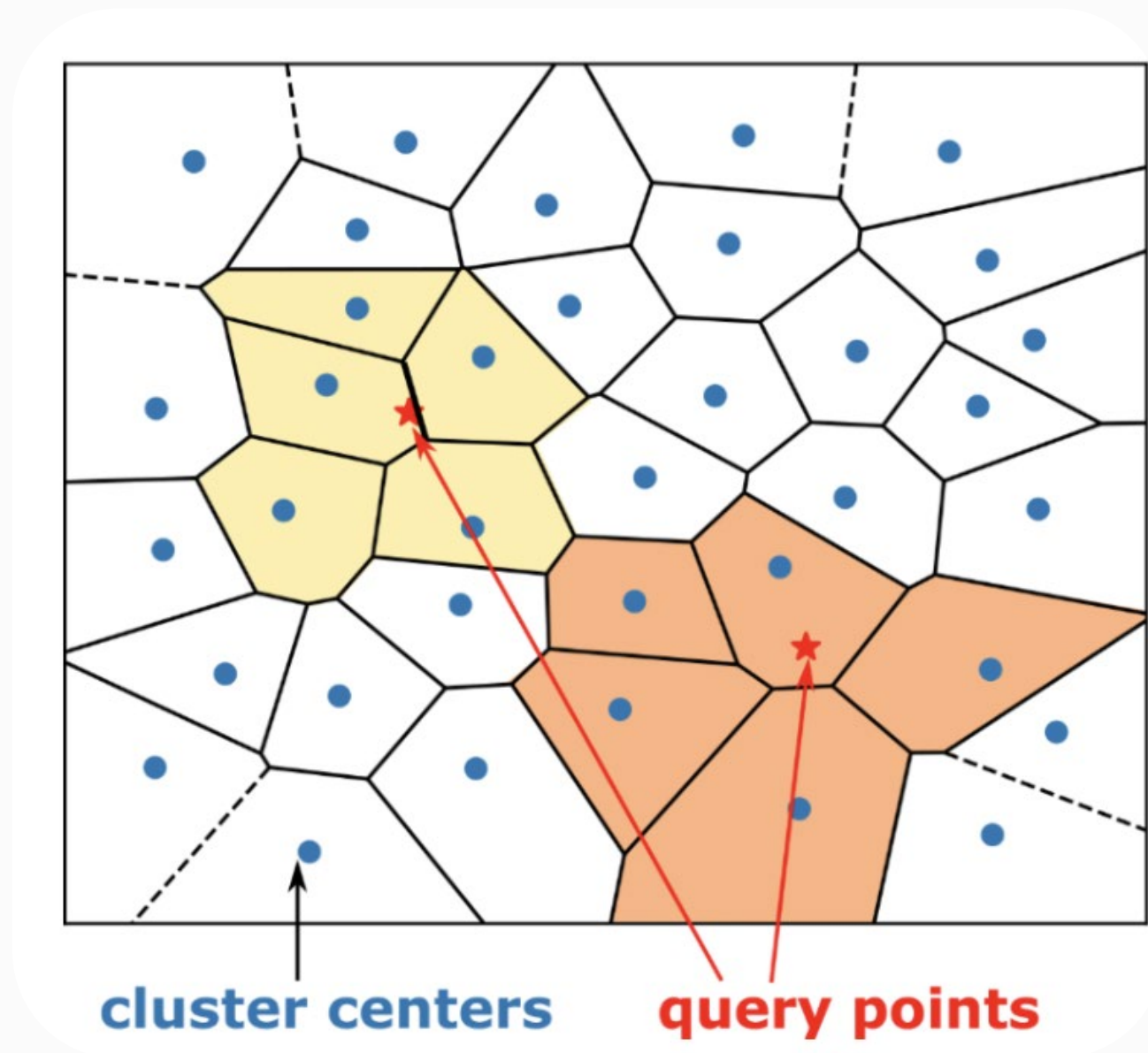
Улучшения классического KNN

- Обычный KNN может работать медленно
- «Ускоряйте» с помощью специальных векторных баз

Векторные базы поиска



Inverted Index



Выводы

- 1 Рассмотрели, как решать задачу рекомендаций с помощью KNN
- 2 Разобрались, какие возможные сложности могут быть в методе коллаборативной фильтрации
- 3 Узнали, что лежит в основе векторных баз данных