

Икрам Магжан
ikmakeg@gmail.com

Коллективный разум в Интернет маркетинге

Содержание

- Что такое коллективный разум?
- Машинное обучение
- Рекомендация товара на основе коллаборативной фильтраций
- Литература

Что такое коллективный разум?

■ Коллективный разум в жизни

■ Задача 1

- Хотим посетить кинотеатр, но не знаем на какой фильм идти лучше
- **Решение:**
 - Спрашиваем у людей, которые посмотрели
 - Рецензия зрителей
 - Оценка кинокритиков

■ Задача 2

- Хотим купить новую художественную литературу, но не знаем какую выбрать
- **Решение:**
 - Смотрим отзывы людей, которые читали эту книжку

- В 2006 году Компания Netflix занимающийся онлайн прокатом DVD, объявляет конкурс на 1 млн. долларов по улучшению точности системы рекомендации фильмов на 10%. 2007 году победителю удалось добиться улучшения на 7%. Пользуясь данными о том, какие фильмы нравятся пользователям, Netflix удается рекомендовать своим клиентам такие фильмы, о которых они даже не слышали.
- В 1998 году запускается новая поисковая машина Google, когда на рынке уже имелось несколько крупных поисковых систем. Основатели разработали новый подход к ранжированию результатов поиска, основанный на использовании ссылок с миллионов сайтов. В 2004 году поисковик обслуживал 85% всех поисковых запросов всего Интернета.
- *Общее между этими компаниями:* выстроили бизнес на применении изощренных алгоритмов объединения данных, полученных от множества людей.

Что такое коллективный разум?

■ Определение

Извлечение нового знания из объединенных предпочтений, поведения и представлений некоторой группы.

Каким образом извлечь полезную информацию?

Машинное обучение или анализ данных(Data Mining)

- Определение
 - подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться.
- Некоторые задачи машинного обучения
 - Задача классификации
 - Задача восстановления регрессии
 - Задача кластеризации

Машинное обучение

■ Задача классификации

- Имеются
 - набор реклам(товаров, новостей и т.д.), которых нужно отобразить на сайте
 - посетители и пользователи этого сайта.
- Какие из реклам нужно показать для определенного пользователя, чтобы тот «кликнул» по ним?
- Для пользователя имеем:
 - Историю продаж
 - Историю посещения страниц

Машинное обучение

- Обучающая выборка
 - История продаж
 - История посещения страниц
- Требуется
 - Построить алгоритм(классификатор), который по набору признаков реклам(товаров, новостей и т.д.) определяет отображать(1-класс) данные на сайте для данного пользователя или нет(2-класс).

Примеры задач классификации

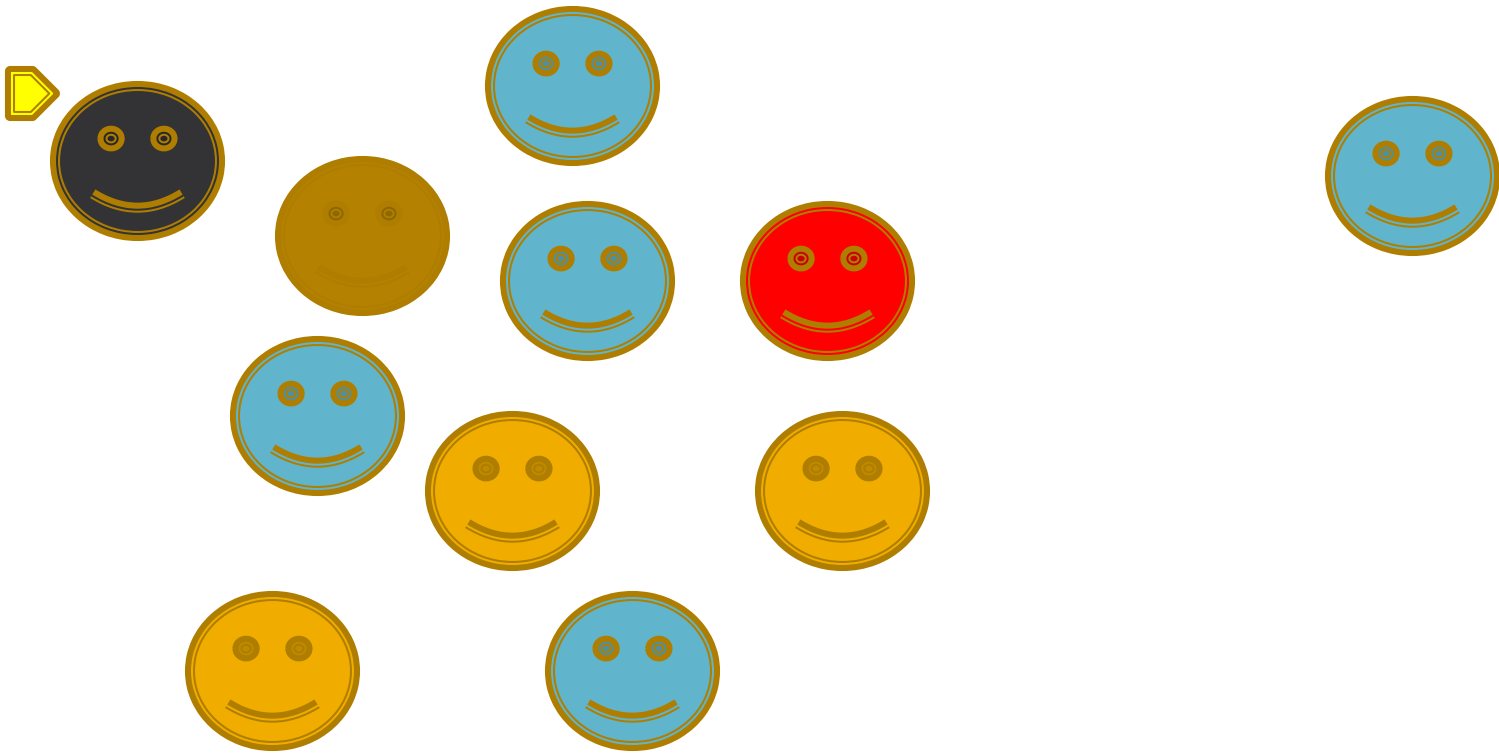
- Медицинская диагностика: по набору медицинских характеристик требуется поставить диагноз
- Геологоразведка: по данным зондирования почв определить наличие полезных ископаемых
- Оптическое распознавание текстов: по отсканированному изображению текста определить цепочку символов, его формирующих
- Кредитный скоринг: по анкете заемщика принять решение о выдаче/отказе кредита
- Синтез химических соединений: по параметрам химических элементов спрогнозировать свойства получаемого соединения

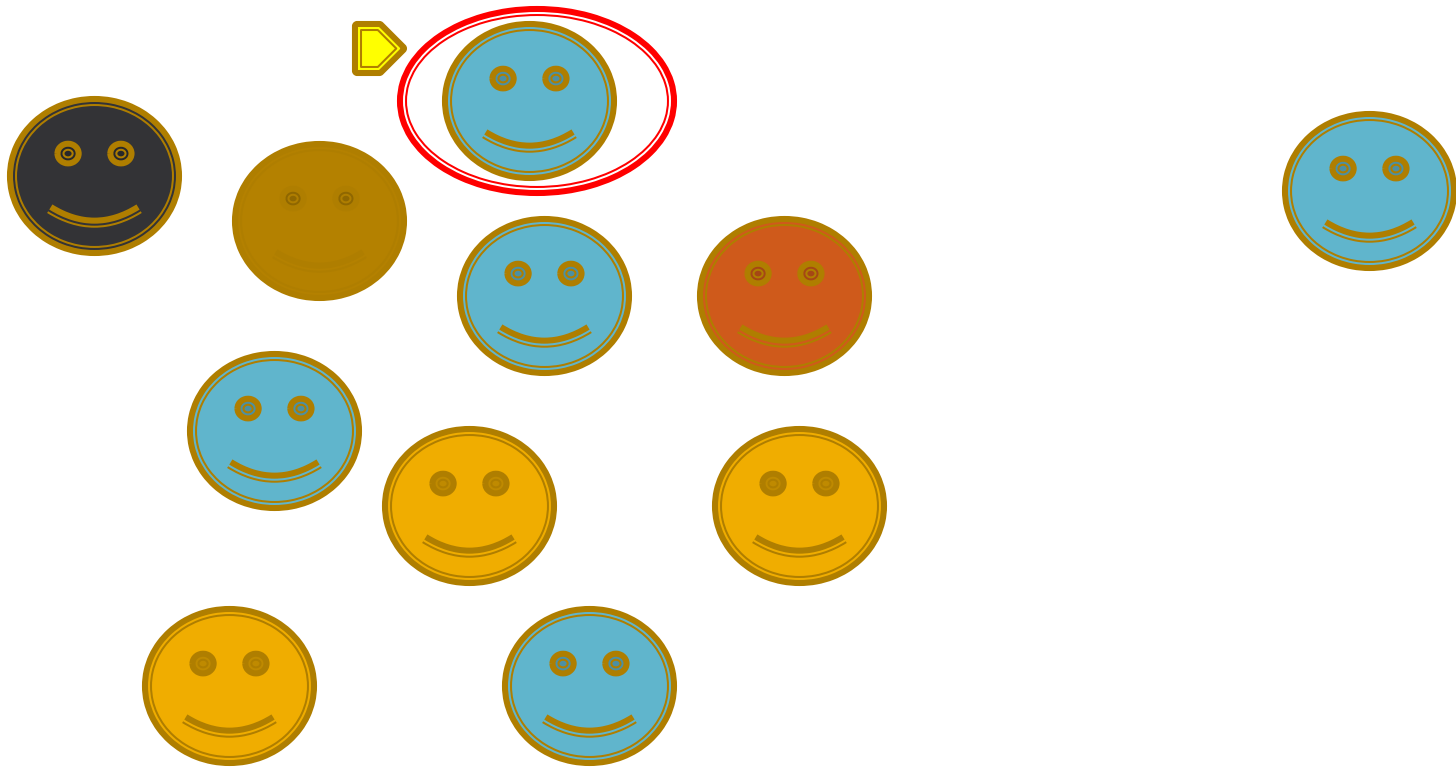
Рекомендация товара на основе коллаборативной фильтрации

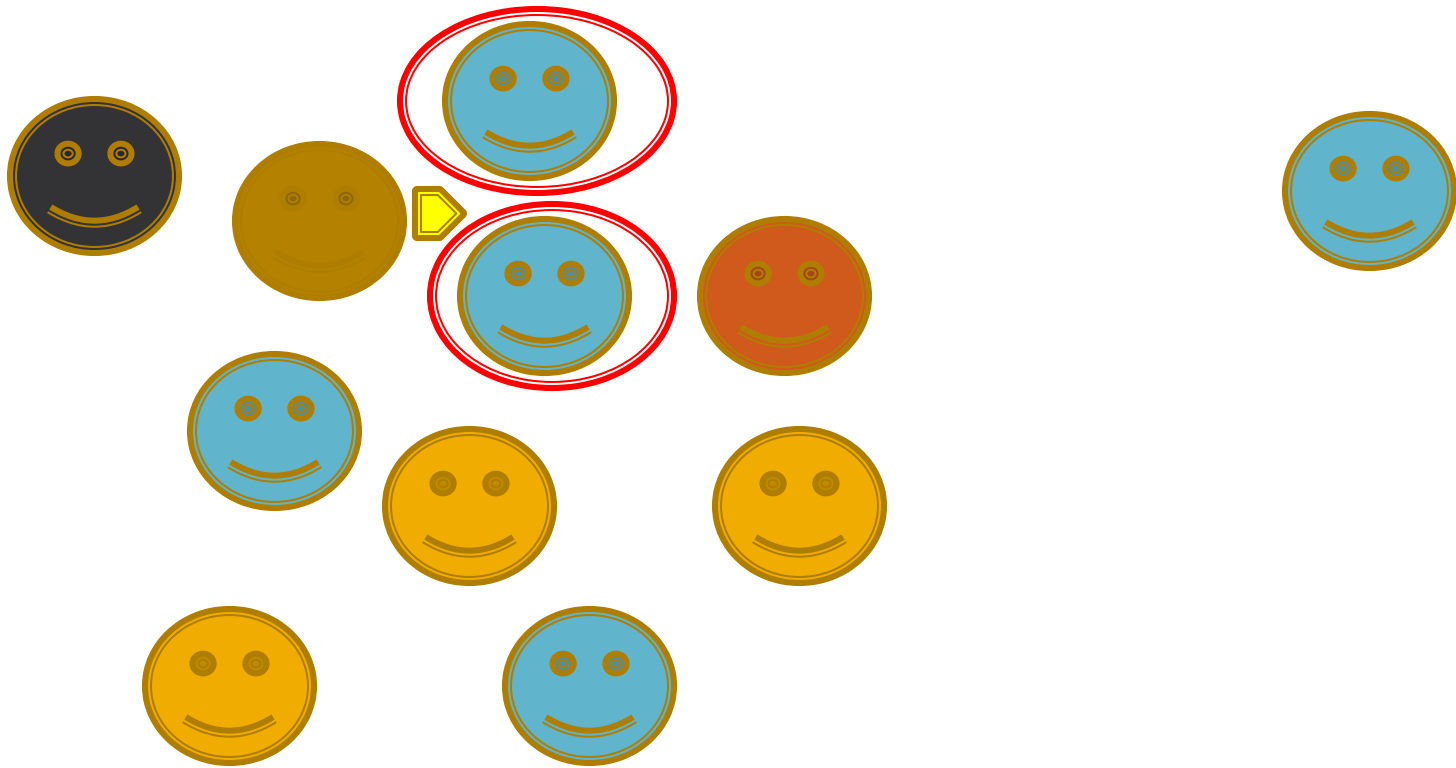
- Рекомендацию можно получить достаточно спросив у друзей, знакомых и т.д.
- Но по мере увеличения количества предложений становится все менее практично основывать решение на основе небольшой группы людей

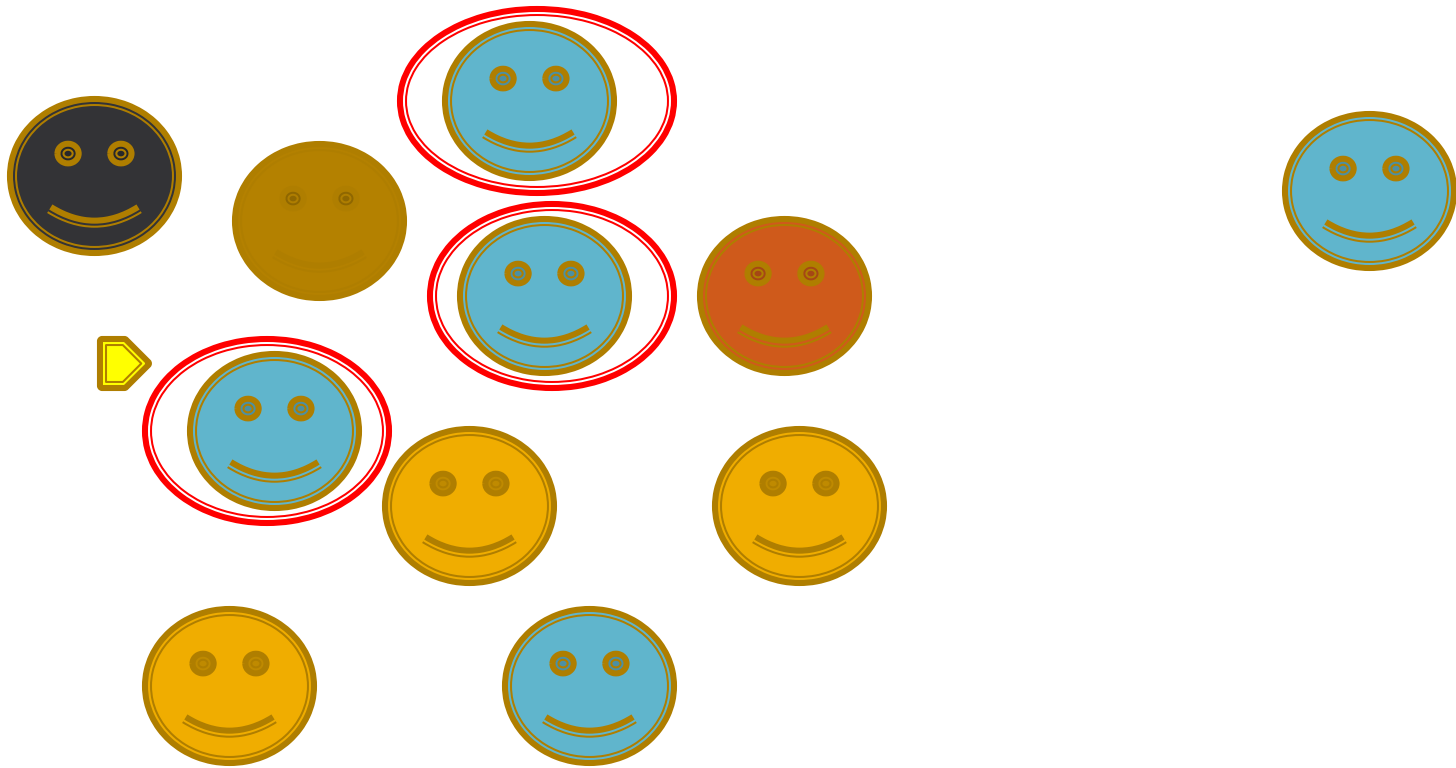
Рекомендация товара на основе коллаборативной фильтрации

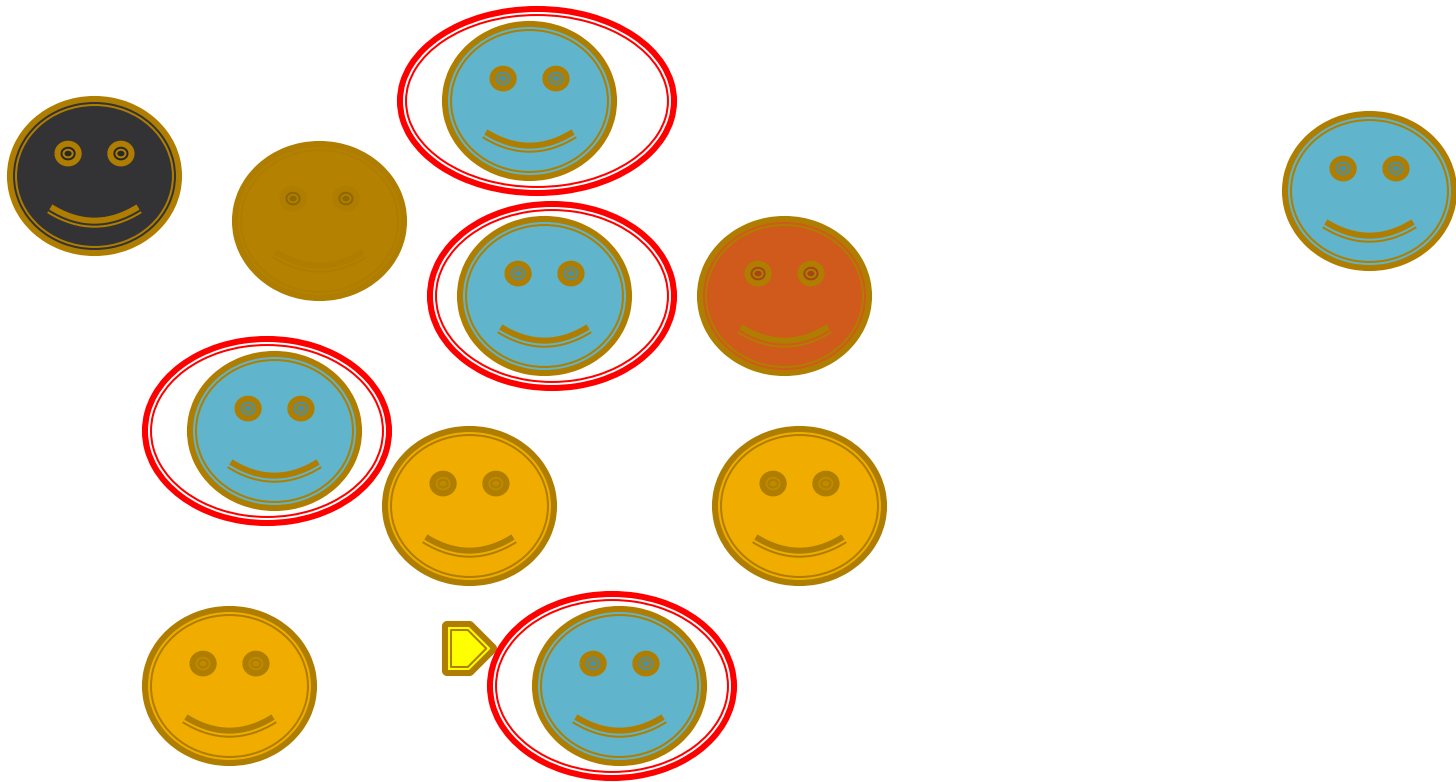
- Коллаборативная фильтрация
 - просматривает большую группу людей и отыскивает в ней меньшую группу с такими же вкусами, как у вас.
 - смотрит, какие еще вещи им нравятся, объединяет предпочтения
 - создает ранжированный список предложений

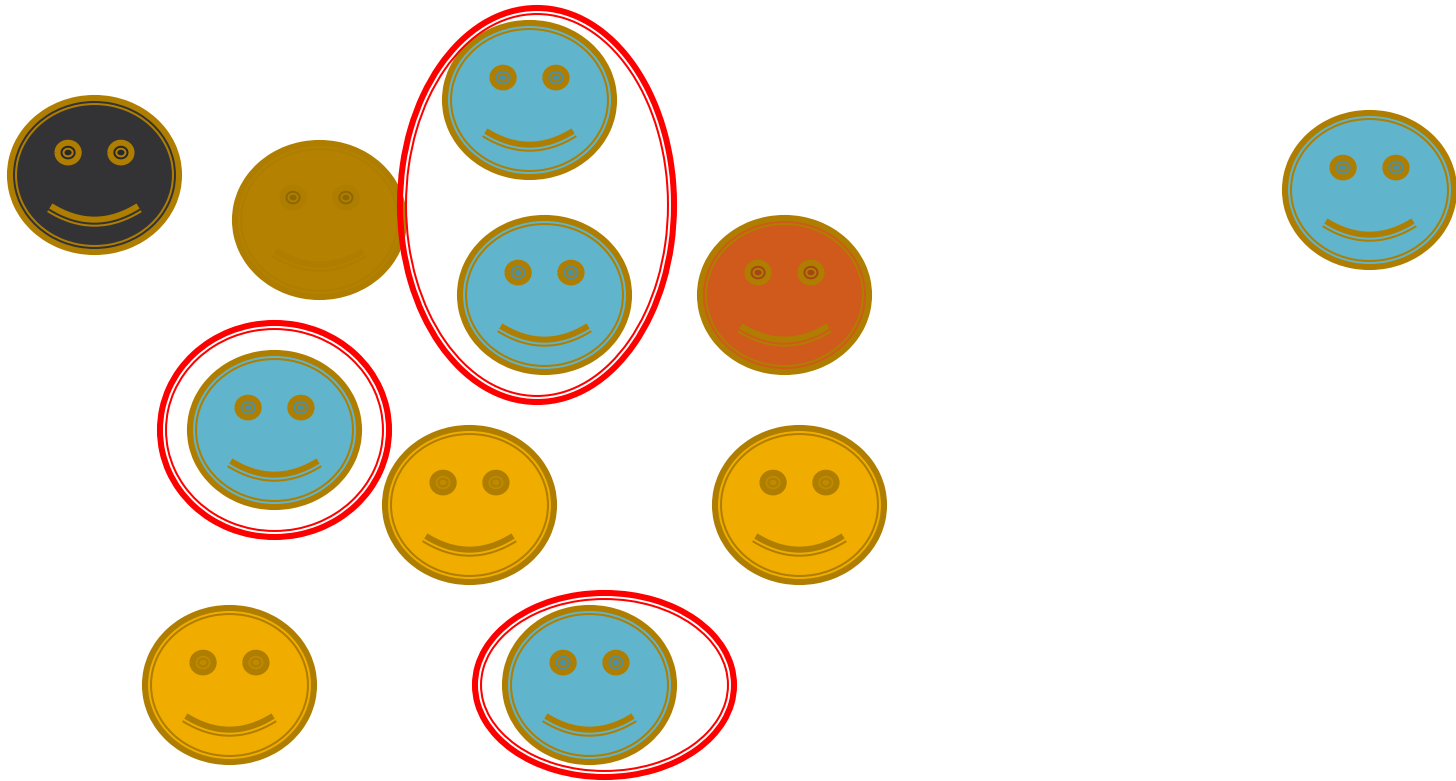












Рекомендация товара на основе коллаборативной фильтрации

- Рассмотрим систему рекомендаций фильмов для определенного пользователя на основе оценок кинокритиков

Рекомендация товара на основе коллаборативной фильтрации

■ Сбор данных о предпочтениях

- Язык программирования Python
- Способ представления кинокритиков – вложенный словарь

	Мстители	Защитник	Жаужүрек мың бала	Мачо и Ботан	Морской бой	Миссия невыполни ма 4
Азиз	3.0	4.0	-	5.0	3.5	3.0
Арман	2.5	3.5	3.0	3.5	2.5	3.0
Юрий	3.0	3.5	1.5	5.0	3.5	3.0
Магжан	-	4.5	-	4.0	1.0	-
Жанар	2.5	3.0	-	3.5	-	4.0
Егор	-	3.5	3.0	4.0	2.5	4.5
Жасулан	3.0	4.0	2.0	3.0	2.0	3.0

Рекомендация товара на основе коллаборативной фильтрации

- Нахождение похожих критиков
 - Собрал данные, нужно определить насколько их вкусы схожи
 - Оценка подобия
 - Коэффициент корреляции Пирсона
 - Евклидовое расстояние
 - Манхэттенское расстояние
 - Метрика Махаланобиса

Рекомендация товара на основе коллаборативной фильтрации

■ Коэффициент корреляции Пирсона

$$r_{xy} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \sum_{i=1}^m (y_i - \bar{y})^2}} = \frac{\text{cov}(x, y)}{\sqrt{s_x^2 s_y^2}},$$

s_x^2, s_y^2 - выборочные дисперсии

\bar{x}, \bar{y} - выборочные средние

$x^m = (x_1, \dots, x_m), y^m = (y_1, \dots, y_m)$; - выборки

$r_{xy} \in [-1, 1]$

Рекомендация товара на основе коллаборативной фильтрации

■ Коэффициенты корреляции Пирсона

-0.258	Жасулан	Жанар
0.396	Арман	Юрий
0.662	Азиз	Магжан
0.991	Арман	Магжан
0.924	Жасулан	Магжан
-1.0	Жанар	Магжан

Рекомендация товара на основе коллаборативной фильтрации

- Ранжирование критиков
 - С помощью оценки подобия ищем наилучшее соответствие между критиками
 - Ищем кинокритиков с таким или наиболее схожим со вкусом пользователя

Рекомендация товара на основе коллаборативной фильтрации

Ранжирование критиков

1.0	Егор
0.99	Арман
0.92	Жасулан
0.66	Азиз
0.38	Юрий

Рекомендация товара на основе коллаборативной фильтрации

- Ранжирование фильмов
 - Вычисляется взвешенная сумма оценок критиков
 - Сортировка по получившимся результатам

Рекомендация товара на основе коллаборативной фильтрации

■ Результат

3.0	Жаужүрек мың бала
3.0	Миссия невыполнима 4
2.5	Мстители

Литература

- Сегаран Т. – Програмируем коллективный разум, O'Reilly, 2008 г.
- Ветров Д. П., Кропотов Д. А. – Байесовские методы машинного обучения, Курс лекции
- Дьяконов А.Г. – Анализ данных, обучение по прецедентам, Издательский отдел факультета ВМК МГУ имени Ломоносова, 2010
- www.machinelearning.ru – Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.

Спасибо за внимание!