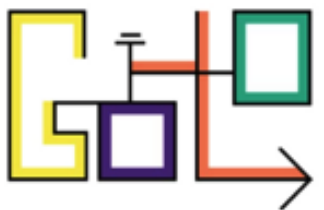


Рекомендательная система для оффлайн- магазина

Александр Данилов



Данные

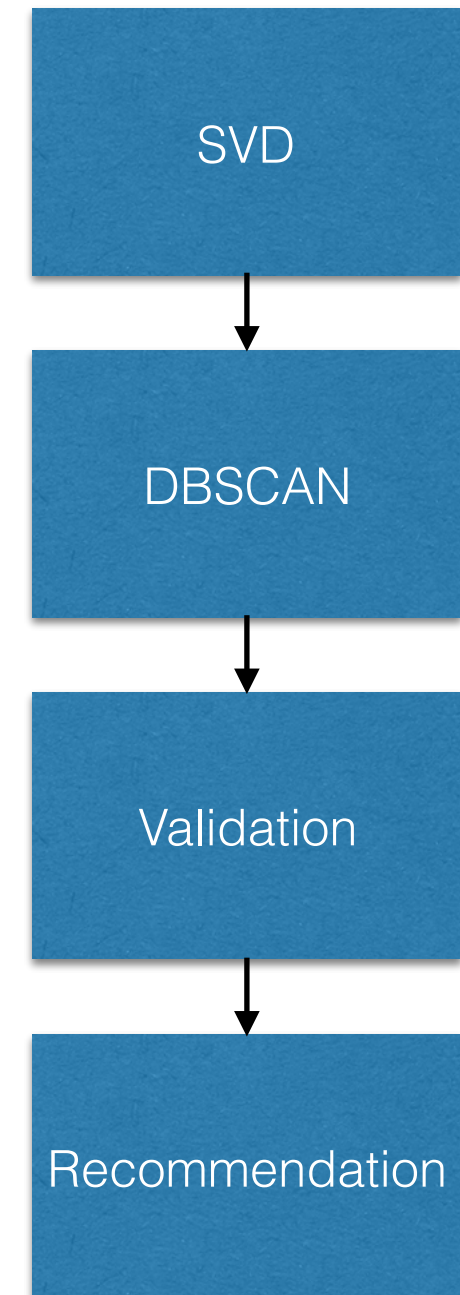
- Чековая лента (дата и время покупки, какие товары купили, какая была скидка)
- Информация о дисконтных картах про 10% чеков
- Информация о товарах

Постановка задачи

Создать рекомендательную систему для
небольшого хозяйственного магазина в
Москве

Что я делал?

1. Создал матрицу “пользователи на товары”
2. Разложил ее при помощи SVD на матрицы “пользователи на признаки” и “товары на признаки”
3. Кластеризовал товары по их признакам с помощью DBSCAN
4. Использовал косинусную меру для оценки кластеризации
5. Для каждого чека определял набор кластеров его товаров и выбирал из всех товаров этих кластеров самые близкие к вектору чека
6. Рекомендовал покупателю топ 5 товаров, самых близких к вектору чека



Метрика

Косинусная мера между векторами чека через кластеры его товаров и через сами товары

$$\cos \angle \left(\overline{clusters}, \overline{items} \right)$$

Результаты

- Качество кластеризации ≈ 0.6
- Адекватные предсказания для некоторых чеков, которые я проверял

Например, для чека с одним товаром:

"СРЕДСТВО Glorix "Лимонная энергия" для мытья пола, бутылка, 1 л, РОССИЯ",

моя система рекомендует товары:

'САЛФЕТКИ Vileda универсальные, п/э упаковка, 3 шт, ГЕРМАНИЯ',

'СРЕДСТВО Domestos "Ультра Белый" универсальное, бутылка, 0,5 л, РОССИЯ',

'ГУБКИ Siral Sissi для мытья посуды, бум.упаковка, 2 шт, ГЕРМАНИЯ',

'ГУБКИ SoftTouch "Двойная сила" для мытья посуды, п/э упаковка, 2 шт, КИТАЙ',

Что можно сделать еще?

- Использовать данные о дисконтных картах, чтобы узнать лучше о конкретном пользователе
- Использовать данные о товарах, чтобы лучше их кластеризовать
- Дообучиться при помощи сервиса, где люди смогут отмечать качество рекомендации (например, бота в Телеграме)

Demo

Спасибо за внимание!

danilov.sasha444@gmail.com