



Факультет компьютерных наук
Образовательная программа
“Прикладная математика и информатика”



Рекомендательная система продуктов онлайн магазинов

ПРОГРАММНЫЙ ПРОЕКТ

Выполнили:
Косимов Аюбджон, БПМИ206

Руководитель:
Паринов Андрей Андреевич

Москва, 2021



Цель: разработка агрегатора товаров с реализацией системы рекомендаций.

Предметная область

- Data Science: обработка, анализ и представление данных в цифровой форме
- Рекомендательные системы: комплекс алгоритмов, задача которого по имеющимся данным выдать список товаров наиболее интересных для конкретного пользователя

Основные термины и определения

- Рекомендательная система
- Content-based filtering
- Collaborative filtering
- Метод сопряженных градиентов
- Проблема холодного старта
- Кластеризация
- Хитрейт

отличается от обычной рекомендательной системы тем, что использует не только контентные, но и социальные связи между пользователями для рекомендаций. Collaborative filtering использует информацию о предпочтениях пользователей, чтобы рекомендовать товары или услуги.

Актуальность

Наличие функционала персональных рекомендаций в онлайн-сервисе как показатель качества.

➤ Яндекс.Маркет

➤ Ozon

➤ Lamoda

➤ Wildberries

➤ e-Katalog

➤ Spotify

➤ Netflix

аналоги

Индивидуальные задачи

	Задача
Газиев С. И.	Реализация алгоритмов рекомендательных систем
Евзман Я. Л.	Создание веб-интерфейса и реализация бизнес-логики
Сармин П. Ф.	Сбор данных с интрнет-магазинов
Косимов А. А.	Организация хранилища данных



Задачи модуля рекомендательных систем

1

Получить общее представление о рекомендательных системах

2

По имеющимся знаниям выбрать алгоритмы для изучения и их последующей реализации

3

Реализовать Content-based filtering рекомендательную систему

4

Реализовать Collaborative filtering рекомендательную систему

5

Реализовать загрузки и обработчики текстовых данных

6

Проверить качество реализованных рекомендательных систем

Функциональные требования

1

Каждый алгоритм рекомендательной системы должен быть обернут в класс, который и будет хранить в себе модель этого алгоритма

2

Для каждой модели должны быть реализованы метод для загрузки данных и метод для построения самой модели

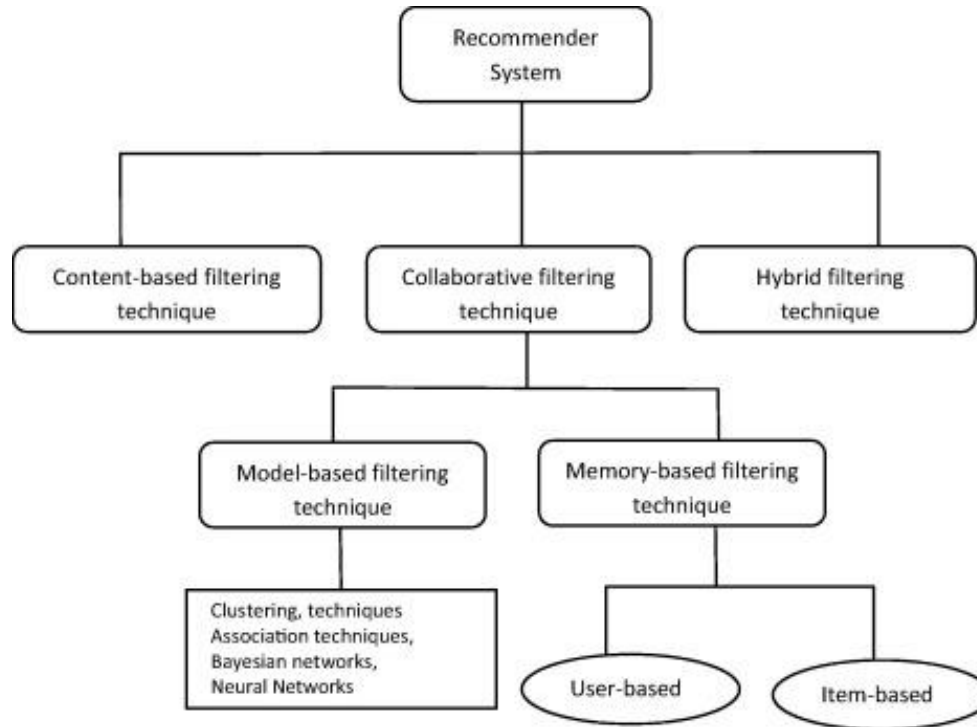
3

Модель должна быть способной возвращать требуемое количество рекомендаций для пользователя

4

Модель должна индексировать товары и пользователей также, как они хранятся в хранилище данных

Анализ существующих подходов



Анализ существующих подходов

	Плюсы	Минусы
Content-based filtering	Достаточно иметь содержимое пользователя или товара	Как правило, работают с текстовыми данными и качество рекомендаций зависит от содержимого
Collaborative filtering	Для модели достаточно иметь данные по отношению пользователя к товарубизнес-логики	Проблема холодного старта; Для хорошей модели нужно много данных о разных пользователях

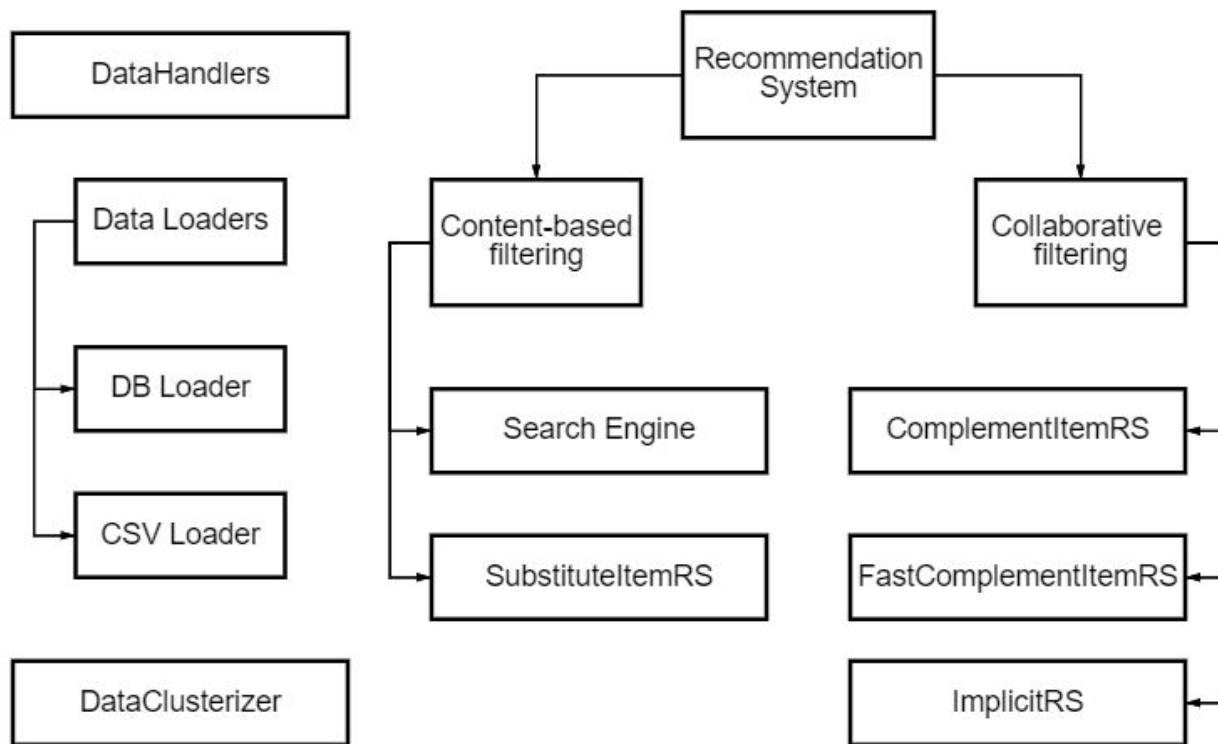
Content-based filtering

- Составление матрицы tf-idf по описаниям товаров (вес слова пропорционален частоте употребления этого слова в тексте и обратно пропорционален частоте употребления слова во всех текстах)
- Результатом запроса query будут индексы товаров, соответствующих наибольшим значениям вектора, полученного при произведении матрицы tf-idf на вектор query.

Collaborative filtering

- Для explicit-модели: получить малоранговую матрицу товаров, используя алгоритм матричного разложения SVD
- Для implicit-модели: разложить матрицу неявных отношений пользователей и товаров с помощью алгоритма ALS и метода сопряженных градиентов
- Для бинарных отношений: ортонормировать матрицу отношений по одной из осей и по полученной матрице сформировать матрицу косинусного сходства по другой оси

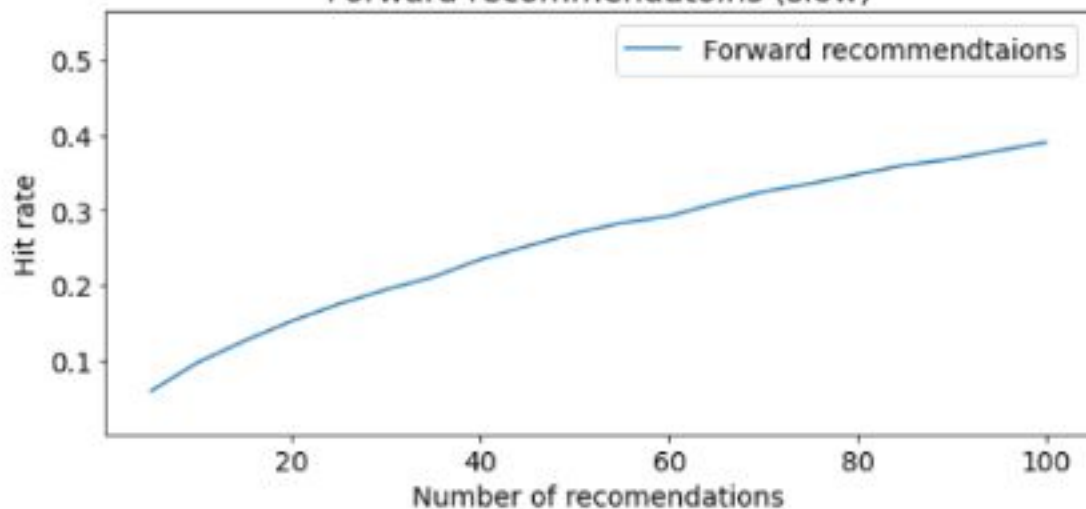
Архитектура модуля



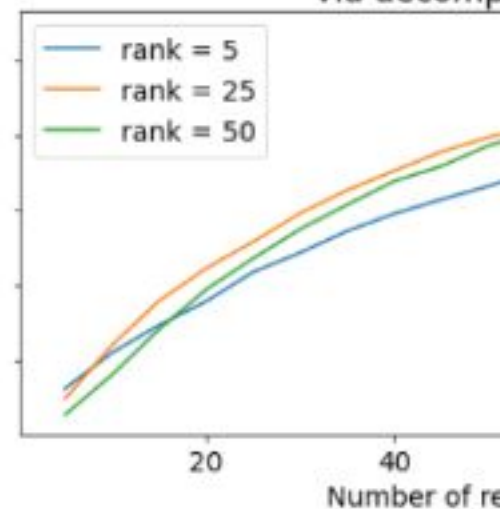
Качество моделей



Forward recommendatoinis (slow)

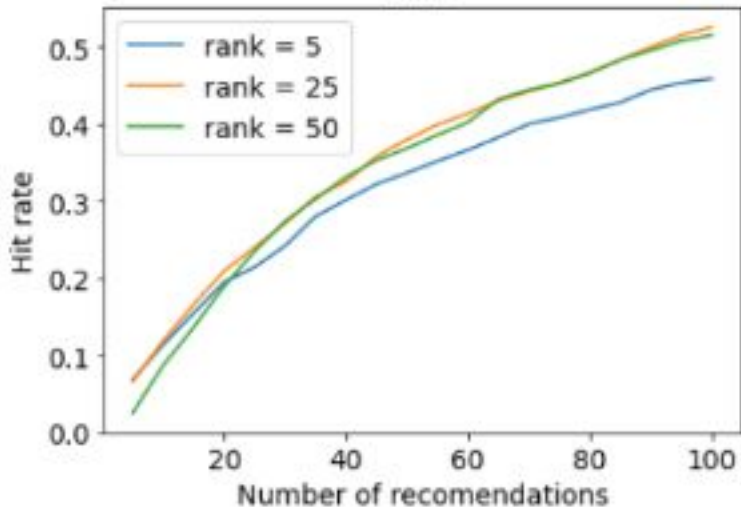


Via decomp

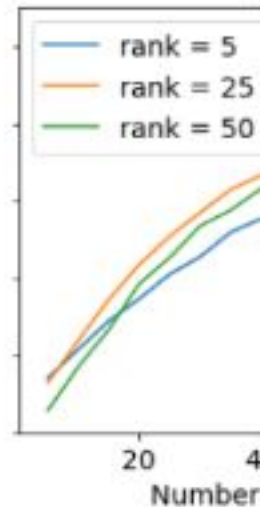
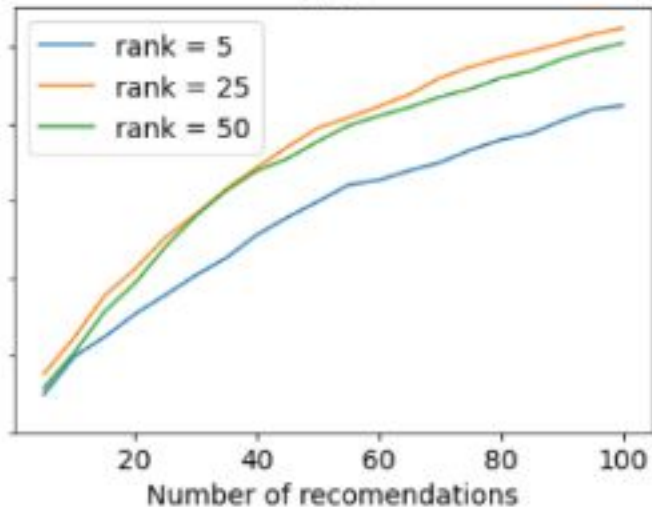


Качество моделей

SVD



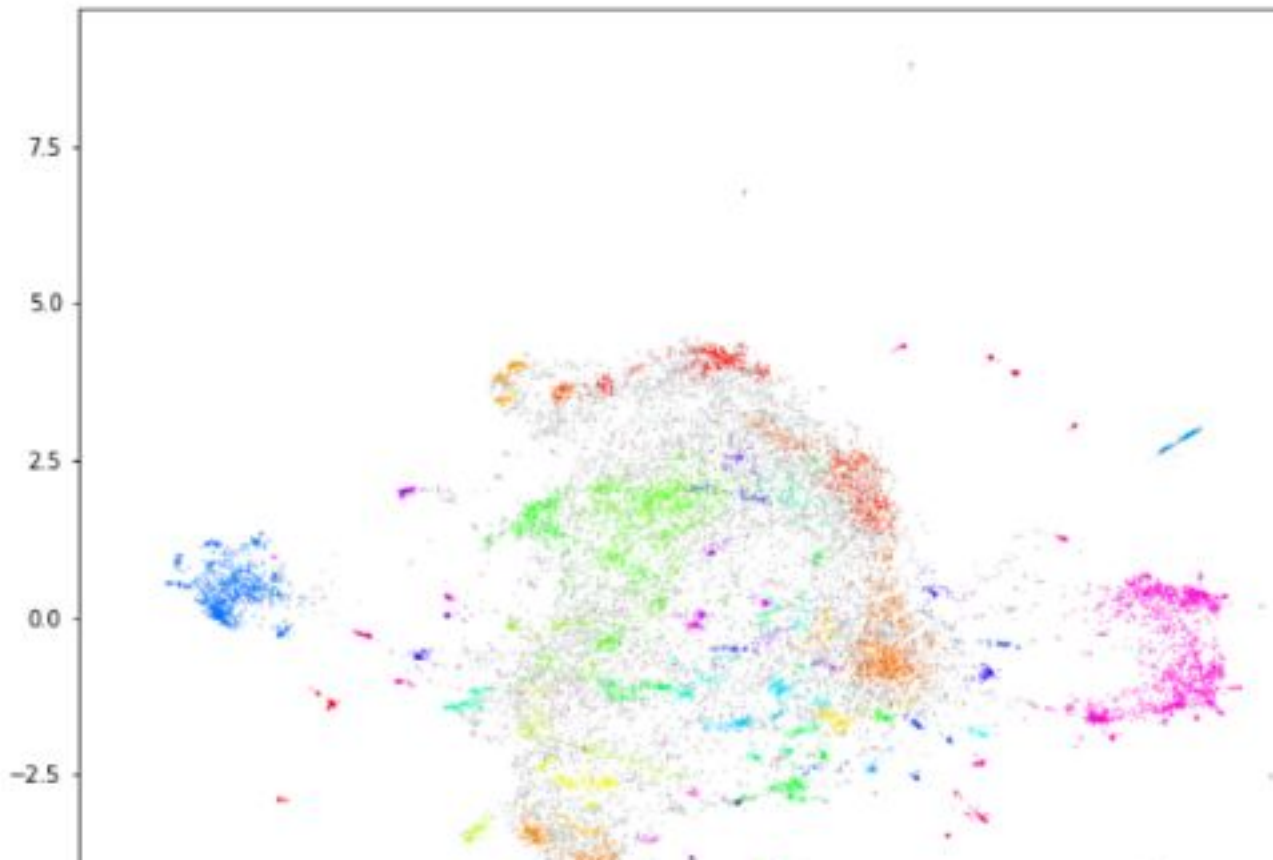
ALS



Кластеризация

- Нужно подбирать количество кластеров
- Приходится обучать модель заново каждый раз, чтобы она адаптировалась под новую выборку данных (по конкретному запросу)

Кластеризация



Задачи были выполнены

- 1 Реализованы 2 Content-based модели
- 2 Реализованы 3 Collaborative-filtering модели
- 3 Реализованы загрузки и обработчики данных
- 4 Проведена проверка качества Collaborative-filtering моделей
- 5 Немного изучил тему кластеризации



Заключение

- Все задачи, поставленные для реализации модуля рекомендательных систем были достигнуты.
- В результате работы над проектом была **достигнута основная цель**: реализован агрегатор товаров с рекомендательными системами.



Направления для доработки

- ▶ расширение списка загрузчиков и обработчиков данных;
- ▶ расширение класса моделей рекомендательных систем;
- ▶ импорт кода на Cython;
- ▶ добавление инструментов для визуализации и анализа данных;

СПИСОК ИСТОЧНИКОВ И ССЫЛОК

1. Git-repository проекта.
<https://github.com/starboy-3/PersonalRecommendationsApp>
2. Git-repository модуля рекомендательных систем.
https://github.com/exponenci/recs_systems
3. Анатомия рекомендательных систем.
<https://habr.com/ru/company/lanit/blog/420499/>
4. Рекомендательные системы: идеи, подходы, задачи.
<https://habr.com/ru/company/jetinfosystems/blog/453792/>
5. F.O.Isinkaye, Y.O.Folajimib, .A.Ojokoh. Recommendation systems: Principles, methods and evaluation.
<https://www.sciencedirect.com/science/article/pii/S1110866515000341>
6. А.Г. Гомзин, А.В. Коршунов. Системы рекомендаций: обзор современных подходов.
https://www.researchgate.net/publication/287544896_Recommender_systems_a_survey_of_modern_approaches
7. Shahzad Qaiser, Ramsha Ali. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents.
https://www.researchgate.net/publication/326425709_Text_Mining_Use_of_TF-IDF_to_Examine_the_Relevance_of_Words_to_Documents

СПИСОК ИСТОЧНИКОВ И ССЫЛОК

8. Dheeraj Bokde, Sheetal Girase, Debajyoti Mukhopadhyay. Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey.
<https://www.sciencedirect.com/science/article/pii/S1877050915007462>
9. Gábor Takács, István Pilászy, Domonkos Tikk. Applications of the Conjugate Gradient Method for Implicit Feedback Collaborative Filtering.
http://www.sze.hu/~gtakacs/download/recsys_2011_draft.pdf
10. Yifan Hu, Yehuda Koren, Chris Volinsky. Collaborative Filtering for Implicit Feedback Datasets.
<http://yifanhu.net/PUB/cf.pdf>
11. Maarten Grootendorst. Topic Modeling with BERT.
<https://www.kdnuggets.com/2020/11/topic-modeling-bert.html>

КОСИМОВ
Аюбджон

группа БПМИ206

aakosimov@edu.hse.ru

Faculty
of
Computer
science
Higher School of Economics