



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени
Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

Отчёт по лабораторной работе №4 по дисциплине "Проектирование рекомендательных систем"

Тема Алгоритмы построения рекомендаций с использованием матричной факторизации

Студент Варламова Е. А.

Группа ИУ7-33М

Оценка (баллы) _____

Преподаватели Быстрицкая А.Ю.

Москва — 2024 г.

СОДЕРЖАНИЕ

Введение	3
1 Аналитический раздел	4
1.1 Коллаборативная фильтрация	4
1.1.1 Коллаборативная фильтрация по пользователю	4
1.2 Матричная факторизация	4
2 Конструкторский раздел	6
2.1 MovieLens Rating	6
3 Технологический раздел	7
3.1 Средства реализации	7
3.2 Библиотеки	7
4 Исследовательский раздел	8
4.1 Условия исследований	8
ЗАКЛЮЧЕНИЕ	10
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	11

ВВЕДЕНИЕ

Цель работы – изучить алгоритмы матричной факторизации на примере SVD применительно к алгоритму коллаборативной фильтрации для построения рекомендаций.

Для достижения поставленной цели потребуется:

- привести описание алгоритма коллаборативной фильтрации по пользователю;
- привести описание алгоритма матричной факторизации;
- привести описание используемых для исследования данных;
- провести исследование скорости работы алгоритма коллаборативной фильтрации и ошибок рекомендаций в зависимости параметров матричной факторизации, сравнить с алгоритмом без матричной факторизации.

1 | Аналитический раздел

1.1 Коллаборативная фильтрация

Коллаборативная фильтрация – это метод рекомендации, при котором анализируется только реакция пользователей на объекты – оценки, которые выставляют пользователи. Оценки могут быть как явными, так и неявными (например, длительность нахождения пользователя на странице товара). Целью фильтрации является предсказание оценки пользователем по оценкам других. Чем больше имеется оценок, тем точнее получатся рекомендации.

1.1.1 Коллаборативная фильтрация по пользователю

В данном методе предполагается, что пользователи, которые в прошлом имели похожие предпочтения, будут иметь похожие предпочтения и в будущем. Для построения рекомендаций с использованием этого метода выполняются следующие шаги:

1. Создание матрицы, где строки представляют пользователей, а столбцы объекты; значения в ячейках матрицы отражают оценки или действия пользователей по отношению к объектам;
2. Вычисление сходства между пользователями (обычно используются косинусное сходство или корреляция Пирсона);
3. Для конкретного пользователя генерируются рекомендации на основе сходства с другими пользователями через агрегацию оценок или действий похожих пользователей.

1.2 Матричная факторизация

Матричная факторизация – это класс алгоритмов коллаборативной фильтрации, используемых в рекомендательных системах. Данные алгоритмы работают путем разложения матрицы взаимодействия пользователя с объектами

на произведение двух прямоугольных матриц меньшей размерности. Зачастую матричная факторизация используется для улучшения качества персонализированных рекомендаций, позволяя выявить скрытые паттерны и взаимосвязи между пользователями и товарами.[1]

Методы матричной факторизации в рекомендательных системах обладают следующими аспектами:

- Снижение размерности – уменьшение объема вычислений и увеличение эффективности;
- Скрытые факторы – данные методы предполагают, что в системе присутствуют некоторые латентные признаки, которые влияют на предпочтения пользователей и характеристики товаров;
- Эффективность работы с разреженными данными – матричная факторизация может эффективно работать с разреженными данными, заполняя недостающие значения.

2 | Конструкторский раздел

2.1 MovieLens Rating

В качестве источника данных был взят датасет, располагающийся в свободном доступе на веб-сайте MovieLens [2]. Набор данных включает в себя множество записей, поля которых описывают идентификатор пользователя, объект, оценку (от 1 до 5) и время появления рецензии. Предобработки проводить не потребовалось.

3 | Технологический раздел

3.1 Средства реализации

В качестве используемого был выбран язык программирования Python [2]. Данный выбор обусловлен следующими факторами:

- Большое количество исчерпывающей документации;
- Широкий выбор доступных библиотек для разработки;
- Простота синтаксиса языка и высокая скорость разработки.

При написании программного продукта использовалась среда разработки Visual Studio Code. Данный выбор обусловлен тем, что данная среда распространяется по свободной лицензии, поставляется для конечного пользователя с открытым исходным кодом, а также имеет большое число расширений, ускоряющих разработку.

3.2 Библиотеки

При анализе и обработке датасета, а также для решения поставленных задач использовались библиотеки:

- pandas;
- numpy;
- matplotlib [3];
- sklearn [4];
- LibRecommender [5].

Данные библиотеки позволили полностью покрыть спектр потребностей при выполнении работы.

4 | Исследовательский раздел

4.1 Условия исследований

Исследование проводилось на персональном вычислительной машине со следующими характеристиками:

- процессор Intel Core i5,
- операционная система MacOS Big Sur
- 8 Гб оперативной памяти.

Временные затраты определялись с использованием библиотеки time.

На рисунке 4.1 представлено исследование алгоритма коллаборативной фильтрации с использованием матричной факторизации в зависимости от количества факторов по времени работы и ошибкам рекомендаций. Кроме того, для сравнения добавлен алгоритм из библиотеки [5], не использующий матричную факторизацию.

Видно, что алгоритм, использующий матричную факторизацию более точен и работает быстрее, чем алгоритм без нее.

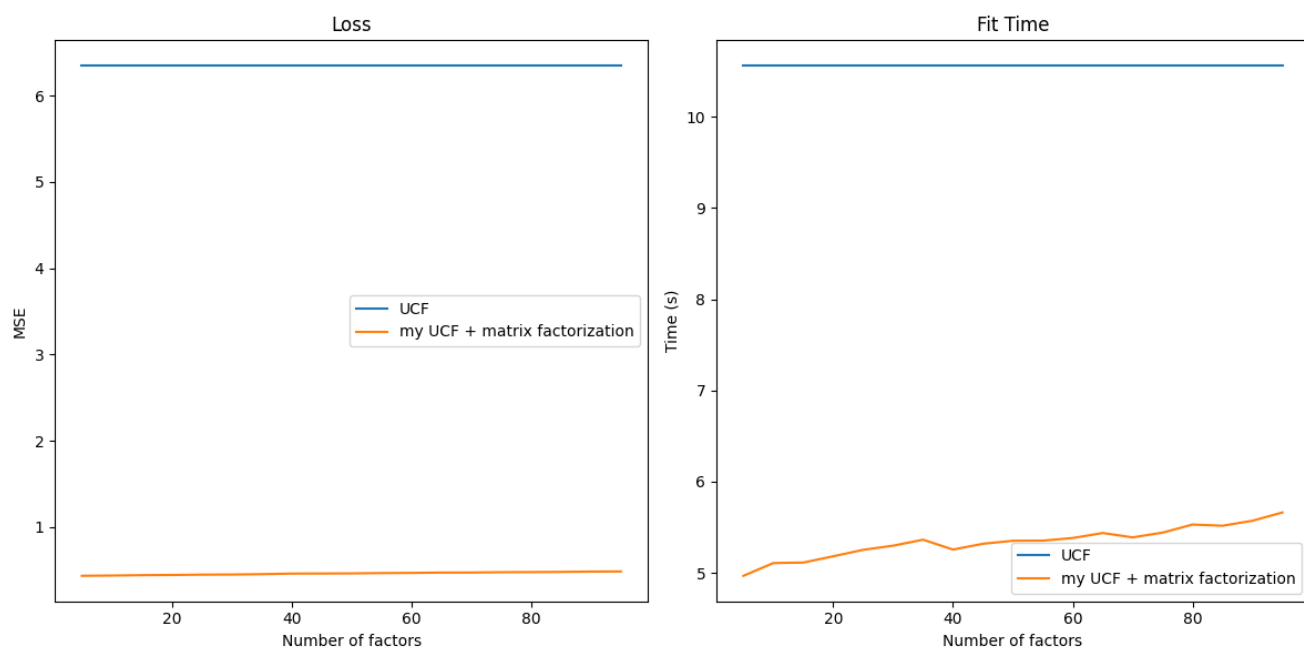


Рис. 4.1: Сравнение алгоритмов

ЗАКЛЮЧЕНИЕ

В ходе выполнения работы были изучены алгоритмы матричной факторизации на примере SVD применительно к алгоритму коллаборативной фильтрации для построения рекомендаций.

Для достижения поставленной цели потребовалось:

- привести описание алгоритма коллаборативной фильтрации по пользователю;
- привести описание алгоритма матричной факторизации;
- привести описание используемых для исследования данных;
- провести исследование скорости работы алгоритма коллаборативной фильтрации и ошибок рекомендаций в зависимости параметров матричной факторизации, сравнить с алгоритмом без матричной факторизации.

Проведенные исследования показали, что алгоритм, использующий матричную факторизацию более точен и работает быстрее, чем алгоритм без нее.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Koren Y. Bell R. V. C. MATRIX FACTORIZATION TECHNIQUES FOR RECOMMENDER SYSTEMS* // IEEE Computer. — 2009. — № 42.
2. Python official page [Электронный ресурс]. — Режим доступа: <https://www.python.org/> (дата обращения 10.10.2024).
3. Matplotlib official page [Электронный ресурс]. — Режим доступа: <https://matplotlib.org/> (дата обращения 10.10.2024).
4. Scikit-learn official page [Электронный ресурс]. — Режим доступа: <https://scikit-learn.org/stable/> (дата обращения 10.10.2024).
5. LibRecommender: official PyPI project page [Электронный ресурс]. — Режим доступа: <https://pypi.org/> (дата обращения 10.10.2024).