

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ, НАУКЕ И ТЕХНИКЕ

УДК 004.942

М. Н. Андреева, Е. Г. Шапович

*Учреждение образования «Барановичский государственный университет»,
Барановичи, Республика Беларусь*

СОЗДАНИЕ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ КНИГ НА ОСНОВЕ ПОЛЬЗОВАТЕЛЬСКИХ ОЦЕНОК

Введение. В современном мире цифровых технологий выбор подходящей книги становится все сложнее из-за огромного количества доступных изданий. Читателям необходимо удобное и эффективное решение, способное предлагать релевантные произведения на основе их интересов и предпочтений. Именно эту задачу решают рекомендательные системы, основанные на анализе данных и пользовательских оценках.

В данной статье мы рассмотрим процесс создания рекомендательной системы книг с использованием языка программирования Python и библиотеки pandas для обработки данных. В основе подхода лежит метод косинусного сходства, который позволяет находить схожие книги, ориентируясь на оценки пользователей. Мы разберем этапы предобработки данных, формирования матрицы взаимодействий и построения модели рекомендаций, которая поможет пользователям открывать новые произведения, соответствующие их вкусу.

Такой алгоритм широко используется в различных сервисах, включая онлайн-магазины и библиотеки, улучшая пользовательский опыт и способствуя популяризации качественной литературы.

Основная часть. Для создания рекомендательной системы мы используем датасет Book Recommendation Dataset, доступный на Kaggle. В этом датасете представлены три файла [1]:

- Books.csv — содержит информацию о названиях книг, авторах и ссылках на изображения обложки.
- Users.csv — включает сведения о пользователях, такие как User-ID, Location, Age.
- Ratings.csv — содержит пользовательские оценки книг.

Этот датасет представляет собой отличный ресурс для построения рекомендательной системы, так как включает реальные пользовательские данные, позволяя анализировать предпочтения и выявлять наиболее популярные произведения.

Каждый файл играет важную роль в рекомендательной системе, так как позволяет связать пользователей, их оценки и книги, создавая алгоритм, который анализирует предпочтения и предлагает похожие произведения.

Перед созданием системы необходимо очистить и подготовить данные:

- проверяем файлы на наличие пропущенных значений и удаляем их;
- удаляем дубликаты, чтобы избежать искажений в рекомендациях;
- объединяем файлы, связывая оценки с книгами и пользователями.

Важно отметить, что многие пользователи выставляют оценки не всем прочитанным книгам, а лишь некоторым. Это означает, что данные будут разреженными, и нужно учитывать этот фактор при построении модели.

Одним из подходов к созданию рекомендаций является анализ популярных книг. Для этого мы:

- считаем количество оценок;
- вычисляем средний рейтинг;
- отбираем книги, получившие не менее 250 оценок, и сортируем их по среднему значению.

Этот метод позволяет выделить топ-50 самых популярных книг, основываясь на реальных данных пользователей.

Для повышения точности рекомендаций важно учитывать только тех пользователей, которые активно оценивают книги. Мы выбираем пользователей, которые оставили более 200 оценок.

Кроме того, отбираем книги, получившие не менее 50 оценок. Это позволяет исключить малозначимые данные и сконцентрироваться на книгах, которые действительно популярны среди активных читателей.

Следующим этапом является формирование матрицы взаимодействий, где строки соответствуют книгам, а столбцы — пользователям. В ней хранятся оценки пользователей, выставленные книгам (рисунок 1).

Такую матрицу удобно использовать для вычисления сходства между книгами, так как она содержит связи между пользователями и их предпочтениями.

В основе системы лежит косинусное сходство, позволяющее сравнивать книги по схожести оценок пользователей. Для поиска похожих книг реализуется функция, которая сортирует книги по степени их сходства и возвращает список наиболее релевантных произведений.

Принцип работы функции:

- находит индекс переданной книги в матрице;
- определяет список схожих книг, сортируя их по степени сходства;
- извлекает название, автора и ссылку на обложку рекомендованных книг.

Пример вызова данной функции представлен на рисунке 2. Мы ввели название книги 1984.

User-ID	254	2276	2766	2977	3363	4017	4385	6251	6323	6543	...	271705	273979	274004	274061	274301	274308	275970	277427	277639	278418	
Book-Title																						
1984	9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1st to Die: A Novel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2nd Chance	0.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4 Blondes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A Bend in the Road	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...
Year of Wonders	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
You Belong To Me	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Zen and the Art of Motorcycle Maintenance: An Inquiry into Values	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Zoya	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
101 Is for Outlaw	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0

Рисунок 1 — Матрица взаимодействий

```

[['Animal Farm',
 'George Orwell',
 'http://images.amazon.com/images/P/0451526341.01.MZZZZZZZ.jpg'],
 ['The Handmaid's Tale',
 'Margaret Atwood',
 'http://images.amazon.com/images/P/0449212602.01.MZZZZZZZ.jpg'],
 ['Brave New World',
 'Aldous Huxley',
 'http://images.amazon.com/images/P/0060809833.01.MZZZZZZZ.jpg'],
 ['The Vampire Lestat (Vampire Chronicles, Book II)',
 'ANNE RICE',
 'http://images.amazon.com/images/P/0345313860.01.MZZZZZZZ.jpg']]

```

Рисунок 2 — Результат работы функции

Как видно из рисунка рекомендации очень даже ожидаемы. На основе книги «1984» Джорджа Оруэлла рекомендательная система подобрала книги в жанре антиутопии, такие как:

- «Скотный двор» — Джордж Оруэлл;
- «Рассказ служанки» — Маргарет Этвуд;
- «О дивный новый мир» — Олдос Хаксли.

Из этого можно сделать вывод, что разработанная рекомендательная система успешно определила тематически схожие книги в жанре антиутопии.

Заключение. Разработка рекомендательной системы книг с использованием пользовательских оценок — это мощный инструмент для персонализированных рекомендаций. В данной статье мы рассмотрели ключевые этапы создания такой системы, начиная с обработки данных и формирования матрицы взаимодействий, до реализации алгоритма на основе косинусного сходства.

Анализ показал, что система успешно определяет схожие произведения, дополняя список книгами с философскими и социальными подтекстами. Это подтверждает эффективность метода, основанного на предпочтениях пользователей.

Рекомендательные системы не только улучшают пользовательский опыт, но и способствуют популяризации литературы, помогая читателям открывать новые книги, соответствующие их вкусам. Внедрение дополнительных алгоритмов и фильтрации по жанрам может еще больше повысить точность рекомендаций, делая их более осмысленными и персонализированными.

Список цитируемых источников

1. Book Recommendation Dataset. — URL: <https://www.kaggle.com/code/azzayahia/book-recommender-system> (date of access: 03.05.2025).