Project Design: Personalized Book Recommendation System

Name: Sihan Zhu

ZID:5536142

1. Project Scope

This project focuses on the field of "online book recommendation" and targets ordinary users and book lovers who use e-book platforms or reading apps. We plan to build a web-based recommendation interface that shows users 5 books at a time. Each book will be accompanied by a cover, title, author name, and several action buttons (such as "want to read", "rate" or "not interested").

Users can express their feedback on the recommendation results by clicking or rating. These feedbacks (whether explicit ratings or implicit behaviors) will be recorded by the system and used to update the recommendation model regularly (for example, every night).

For new users, we will use a content-based recommendation method to generate recommendations based on the book's meta information (type, author, publication time, etc.); we will also use the "hot books" on the platform as initial recommendations to help them find content of interest faster.

From a business perspective, such a recommendation system can improve user retention, bring book sales revenue through promotional links, and even provide publishers with valuable user reading preference data (anonymous of course).

2. Dataset

We will use the Book Recommendation Dataset available on Kaggle, which contains three main files:

Books.csv: contains the title, author, publisher, and year of publication of the book;
Users.csv: contains the user's location and age information;
Ratings.csv: records the user's rating information for the book (both explicit and implicit).

Our recommendation model will mainly build a user-item rating matrix based on Ratings.csv, and then combine Books.csv for content-based recommendations. The user features in Users.csv will also help us achieve more personalized recommendations (such as user clustering).

Of course, the dataset also has limitations, such as lack of time information and incomplete user information, but it contains hundreds of thousands of interaction data, which is large enough to support the development of an effective recommendation system.

3. Recommendation Methods

We plan to build a hybrid recommendation system, which includes the following methods:
User-based collaborative filtering (such as k-NN): find similar users based on user ratings;
Content-based filtering: use book meta information to recommend similar books;
Matrix decomposition methods (such as SVD or ALS): mine potential features through rating matrices.
In addition, we will try different combinations, such as comparing the effects of "user-based" and "item-based" collaborative filtering, or using weighted ensemble models to fuse multiple recommendation results. These methods can be implemented using common Python libraries such as surprise, LightFM or scikit-learn, adapted to the data format provided by Kaggle.

4. Model evaluation and user research (Evaluation)

We will use the following indicators to evaluate the recommendation effect:
Offline evaluation (based on historical rating data):
Precision, recall, F1 score (Top-N recommendation)
nDCG@N: Consider the quality of recommendation ranking
RMSE/MAE: measure the accuracy of predicted ratings
System-level evaluation (based on user interaction data):
Click-through rate (CTR): measure whether the recommendation attracts users
Diversity/novelty: avoid recommending duplicate and too popular content
Coverage: whether the system can recommend a wide range of books
Computational performance:
The model will be trained offline regularly, and the online query response time must be controlled within 100 milliseconds;
The model is updated once a week and supports incremental training or fine-tuning.
User research design:
Plan to find 10 students to participate in the simulation test and use web pages for interaction;
Collect interaction logs + design a short questionnaire (based on Likert scale) to obtain feedback on recommendation relevance, satisfaction, etc.
During the research process, we will ensure that all data is anonymous and no sensitive information will be collected or used.

Summary

This project uses a practical and popular recommendation field (books), and the data is reliable and large enough to make a complete and scalable recommendation system. We hope to achieve good results not only in terms of model accuracy, but also in terms of user feedback to make a practical system. The project is highly feasible and suitable for completion within the time frame of this semester.