Collaborative Filtering Report

1. Introduction

This report presents a comparative analysis of collaborative filtering techniques applied to the Book-Crossing dataset. The objective is to design an effective recommender system by leveraging:

- User-User Collaborative Filtering
- Item-Item Collaborative Filtering
- Matrix Factorization using Truncated Singular Value Decomposition (SVD)

Each technique is implemented, evaluated, and compared based on its recommendation quality and suitability under different scenarios.

2. User-User Collaborative Filtering

Methodology

User-User collaborative filtering computes cosine similarity between user vectors in the interaction matrix. For a given user, it identifies the K most similar users and recommends books they have rated highly but the target user has not rated.

Advantages

- Captures personalized preferences through peer similarity.
- Simple to interpret and explain to end-users.

Limitations

- Performance degrades for users with few interactions (cold-start problem).
- Requires substantial overlap in rated items to identify similarity effectively.

3. Item-Item Collaborative Filtering

Methodology

This approach calculates cosine similarity between item (book) vectors. For each book rated highly by a user, the system finds similar books and recommends them if the user has not interacted with them.

Advantages

- More stable than user-based methods in sparse datasets.
- Performs well when user interactions are limited but involve popular items.

Limitations

- Ineffective when rated books are rare or lack sufficient co-occurrence.
- May return generic recommendations with limited diversity.

4. Matrix Factorization (SVD)

Methodology

Matrix Factorization via Truncated SVD reduces the high-dimensional user-item interaction matrix into latent feature space. This allows for the reconstruction of missing ratings using learned user and item profiles.

Advantages

- Effectively handles sparsity by learning latent relationships.
- Generates recommendations for users and items with minimal data.

Limitations

- Less interpretable than neighborhood-based methods.
- Requires careful selection of the number of latent components.

5. Evaluation and Observations

Experimental Observations

- All three models were evaluated on a subset of users.
- User-User filtering performed well for users with rich interaction histories.
- Item-Item filtering returned high-quality recommendations when users rated popular books.
- SVD proved to be the most consistent, particularly in sparse and cold-start situations.

Case Example (User ID: 626)

- No recommendations were generated by the Item-Item model.
- SVD successfully predicted relevant books, highlighting its robustness in sparse scenarios.