Book Recommendation System

Anupam Garg
anupam20555@iiitd.ac.in

Saharsh Dev
saharsh20572@iiitd.ac.in

Saksham Singhi
saksham20463@iiitd.ac.in

Vishal Kumar
vishal20154@iiitd.ac.in

Vanisha Singh
vanisha20347@iiitd.ac.in

Vanshika Goel
vanshika20413@iiitd.ac.in

Abstract

With the rise of many social media platforms Like Amazon, Netflix, Youtube, recommender, systems have taken more and more place in our lives.From e-commerce (suggest to buyers articles that could interest them) to online ads recommendation systems are today unavoidable in our online journeys.

Concept
Book Recommendations, Content-based filtering

Motivation
Recommender systems are the algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy, or anything else depending on industries). Such systems are really important in some industries. It can generate a huge amount of income and this is a way to stand out significantly from competitors.The main objective is to create a book recommendation system for users.

Updated Problem Statement
With millions of books available, it may be little difficult for people to select and choose what to read next, and sometimes, the contents of the books don't turn out what they were expected to be. A book recommendation system will personalize, increase efficiency and engagement and diversify the range of books.

Literature Review

1. In the Book Recommendation System[1], the authors proposed three major approaches for the recommendation systems, i.e content-based, collaborative and hybrid. The systems which implement the Content-based approach recommend items to a user which are similar to a user which are similar to the one the user preferred in the past. The recommender system that implements the Collaborative Filtering predicts the users' preference by considering the similarities between them. Hybrid approaches tries to combine the content and collaborative based approaches, which have both positive and negative sides, thus produces better results.

2. In the paper Book Recommendation System[2] the authors discussed the various approaches which is used to design such recommender systems. The content based filtering recommends new entities by using more custom data to retrieve the best matches. Similarly, the collaborative filtering requires large dataset of active users who have previously evaluated the product to make the correct prediction. But both of these approaches had limitations & they suggested to combine the two systems to form a hybrid system and allows to combine item properties with other user's preferences. The recommendation engine employs a collaborative social networking approach which tries to mix the taste with a large audience to produce meaningful results.

3. In the paper[3], the authors explained a clustering-based book recommendation system that uses different approaches, including Collaborative, hybrid, content-based, knowledge based, and utility-based filtering. Clustering allows regrouping all books based on the rating and user preference datasets. Such clustering shows remarkable prediction capability for a personalized book recommendation system.

Baseline Results

Before updating the baseline, it is important to note that the raw book ratings are obtained after merging the data of all the users, books and the book ratings. Earlier, we had used the Model-Based Approach of collaborative Filtering to get appropriate recommendations for the user. After fitting the data into a useful model(K-NN), we obtain our predicted ratings and used the Root-Mean Square Error(RMSE) to calculate the loss/error of the prediction ratings. It tells on an average how different are our predictions from the actual ratings which the user has provided. In this case, it was found to be about 4.9537.

In the updated version, we have the same books-Crossing (BX) dataset which has three tables for the Books, Users and the Ratings respectively. We have implemented a couple of algorithms over building and coming up with a conclusion of our proposed method.
Some of the algorithms which were implemented are:

1. Popularity - based approach: We have used a Popularity-based approach which works on the rules of popularity and looks for those books which are trendy among the users. This is relevant to prevent Cold-Start Problem which tries to draw the issue of unavailability of any inference/recommendation for a new user/item about which it hasn't collected much information.
The negative part is that it is not sensitive to the interest and tastes of a particular user and is not personalized.

2. User-Item Collaborative recommendation: This works by considering the user-ratings and finding the cosine similarities in the ratings by many users to recommend books. We considered only those books which have been rated by at least 50 users.

```
Input Book:
Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))

RECOMMENDATIONS:

Harry Potter and the Prisoner of Azkaban (Book 3)
Harry Potter and the Goblet of Fire (Book 4)
Harry Potter and the Chamber of Secrets (Book 2)
Fried Green Tomatoes at the Whistle Stop Cafe
Cold Sassy Tree
```

3. Correlation-based Recommendation: We have created the correlation matrix which considers only those books which have a total rating of more than 50. Then a user-book rating matrix is created. For the input book, top-books which are highly co-related to it are recommended.

ISBN	Book-Rating	ratingCount
0971880107	4.387665	2502
0316666343	8.226040	1295
0385504209	8.384615	883
0060928336	7.864000	732
0312195516	8.189781	723

4. Content-based approach: This type of system recommend books by calculating the similarity in the book-titles. The TF-IDF feature vectors were created for the unigrams and the bigrams of titles.

```
Recommended Books:

A Bend in the Road
Tara Road
Dreamcatcher
Bridget Jones's Diary
White Oleander : A Novel (Oprah's Book Club)
```

Here, we divided our data based on the explicit ratings provided. It refers to picking up the non-zero rating values from dataset and splitting it into the training and the testing sets.

```
## Explicit Ratings Dataset
from sklearn.model_selection import train_test_split
dataset1 = dataset[dataset['Book-Rating'] != 0]
dataset1 = dataset1.reset_index(drop = True)
dataset1.shape

trainset, testset = train_test_split(dataset1, test_size=.25)
print(trainset.shape)

(287881, 9)
```

```
Input Book:

Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))

Recommended Books:

Harry Potter and the Prisoner of Azkaban (Book 3)
The Street Lawyer
Harry Potter and the Goblet of Fire (Book 4)
Insomnia
Harry Potter and the Order of the Phoenix (Book 5)
```
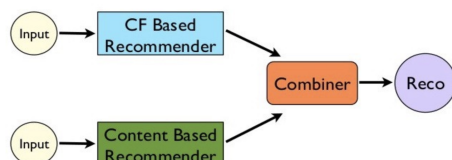
On using the average ratings as calculated while determining the book-ratings using Correlation-approach, we used our training data to fit the ratings to average ratings dataframe. On calculating the RMSE value,we obtain error as 2.5124, which is better as compared to our initial baseline performance. This shows that our system has improved its recommendations as the ratings are in close proximity to the actual ratings provided by the user.

Proposed Method:
We propose a hybrid-recommendation system by using the combination of both content-based and collaborative based filtering systems. It tries to reduce some of the issues that each system encounters individually, like sparsity of matrices etc. Our hybrid system would work by weighing the probability of what a user would think about the goodness of the two systems related to a book.

We can make use of a percentile score to the results obtained from the content and collaborative based filtering models and combine them to recommend the top n-books.

Data - Feature and Analysis
We can obtain the book features/genres from the tags, which are provided by the user. For each book, we would get its tags and it will contain the genre of the book. For this,we may have to look at some other Datasets for a better visualization of the proposed solution.
To add some innovation to our Recsys, we would like to implement some more advanced algorithms like neural networks/deep learning which can give more accurate and personalized recommendations.

References
[1]Cho Jinny, Gorey Ryan, Serrano Sofia et al. 2016 Book Recommendation System, Advisor: Prof. Anna Rafferty, Carleton University, Canada

[2] Mishra Himanshu, Asthana Ashish, Book Recommendation System 2022 IJERT (International Research Journal of Engineering and Technology), Page 2953 - 2959

[3] Sarma Dhiman, Mittra Tanni, Hossain Shahadat Mohammad, Personalized Book Recommendation using Machine Learning algorithms, IJACSA (International Journal of Advanced Computer Science and Applications) 2021.

Hybrid Recommendations