



BOOK RECOMMENDATION SYSTEM

¹N.JayaLakshmi, ²E.J.Srujana, ³K.Keerthi, ⁴K.Jyothi, ⁵K.Manoj, ⁶P.Shankar Rao

¹ Assistant Professor, Department of CSE, Raghu Institute of Technology, Visakhapatnam, Andhra Pradesh, India.

^{2,3,4,5} Student, Department of CSE, Raghu Institute of Technology, Visakhapatnam, Andhra Pradesh, India.

Abstract: When the number of books provided by the existing system is relatively large, it becomes difficult for users to select appropriate books from a lot of books. So, our goal is to develop a web application that recommends books based on users ratings. To recommend books we are using machine learning algorithm like collaborative filtering[CF] that first construct the user-item interaction matrix, then construct vector matrix using cosine similarity measure from user-item interaction matrix and then find the similarity between the books using vector matrix and recommend the top n books similar to the book given by the user as input to the algorithm. By default, our application recommends the books based on books popularity and Users can also search for the books of authors or publishers by giving one of the books of the author or publisher as input. The evaluation metric Root Mean Square Error(RMSE) is used to evaluate the performance of the recommender system. The results indicate that recommendation performance is better when cosine vector similarity measure is used.

Keywords: Collaborative Filtering Technique[CF], Cosine Similarity, Euclidean distance similarity, RMSE

1. Introduction:

Recommender systems are software tools and techniques which provide suggestions to the user for the item or services useful for them. Recommender systems perform an important role in finding the users interests. When we want to read a new book we generally ask our friends or others. After all asking and searching, we may still not find any book of our preference as not everyone has the same interests. For such situations, we need a system which takes our choices into consideration and suggests to us some good books. Today's internet users don't want to spend much time looking for a particular item of their interest. So, this paper proposes a web application where users can give the name of the book as input and top-n rated items similar to the input item are recommended. For recommendation, we implemented a collaborative filtering algorithm with cosine similarity metric for measuring the similarity between the books. By default, our application recommends the books based on books popularity i.e. by calculating weighted-average score for each book and giving top n-scored books as output. Users can also search for the books by the author or publisher. If the user gives the book name and searches for the same author books or publisher books, our application will extract the author or publisher of the given book from the dataset and recommend other top-rating books by him/her. This paper also proves that the performance measure of the algorithm is better when cosine similarity metric is used rather than euclidean distance. The application behind the book recommendation system is also used in libraries and e-commerce websites like kindle, amazon, youTube and even in some book stores.

2.Related Work:

Recommender systems have become extremely common, and are utilized in a variety of areas: some popular applications include movies, books, research articles, and social tags. There are three basic categories of recommendation algorithms: collaborative filtering, content based filtering, and hybrid recommendation.

Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users[2]. The proposed algorithm does not need professional knowledge, and the recommendation effect will become better and better with the interest of the user, but there are data sparsity and other problems. There are two types of Collaborative Filtering techniques: Item-based[7] and User-based[4] collaborative filtering[CF]. Similarity between items or users can be calculated using different similarity metrics like cosine, euclidean, jaccard[1]

Content-based filtering[7][8] methods are based on a description of the item and a profile of the user's preference. These algorithms try to recommend items that are similar to those that a user liked in the past. Hybrid recommendation[2][8] is combining collaborative filtering and content based filtering. These methods can also be used to overcome some of the common problems in recommender systems such as cold start and the sparsity problem[5].

Most researchers used the Cosine Vector Similarity function to compute similarity among book ratings to recommend books, to take into account that users have different rating schemes. Some rate items high usually while others usually rate low[1].

3.Data Collection:

Dataset plays a major role as it is given as input to the machine learning model and output is predicted based on the data in the dataset. Datasets were collected from the Goodreadsbooks repository of kaggle which are of ".csv" format i.e. Books.csv, Ratings.csv, Users.csv.

The description of datasets are as follows:

- Books.csv: Attributes are ISBN, Book-Title, Book-Author, Year-Of-Publication, Publisher, Image-URL-S, Image-URL-M, Image-URL-L
- Ratings.csv: Attributes are User-ID, ISBN, Book-Rating
- Users.csv: Attributes are User-ID, Location, Age

4.Data Cleaning:

As the data is in textual content format, unstructured textual data need to be preprocessed to make the dataset standardized and noise-free. So datasets need to be cleaned by way of checking for null values and delete those rows from the dataset, and also check if any string exists in integer data type or for any misplaced values and modify those rows. This can be done using the pandas library. After all three datasets are standardized individually, all three need to be merged as one dataset based on unique columns to feed to the algorithm.

5. Proposed Methodology:

Collaborative Filtering Technique:

Collaborative filtering[CF] technique focuses on the relationship between users and items. It is a technique that can filter out items that a user might like on the basis of ratings given by the other users and recommends the top-n similar items to the item given by the user. The similarity of items is determined by the similarity of the ratings of those items by the users who have rated both items.

5.1: Item-based Collaborative Filtering Technique:

Item-based collaborative filtering was developed by Amazon. In an existing system where there are more users than items, item-based filtering is faster and more stable than user-based. This algorithm uses a similarity measure to find similarity between items.

Steps involved in this algorithm:

Step 1: Develop a user-item interaction matrix from the merged dataset as shown in the fig.1 and generate vectors from it. Here the item refers to a book.

Step 2: Now construct a cosine matrix using cosine similarity from vectors.

Step 3: From cosine matrix find the books similar to the book given by the user

Step 4: Recommend the top 'n' similar books

Suppose we have n users set[7] $U = \{user1, user2, \dots, user_n\}$ and m Books set

$B = \{book1, book2, \dots, book_m\}$.

User-item matrix is $m \times n$ matrix and values will be the ratings. Vectors will be the book1, book2, ..., book m. Similarity between the books is calculated.

5.1.1 Cosine Vector Similarity(CVS):

Cosine Vector Similarity[2] is the dot product of the two data points(vectors). It measures the cosine angle between the objects i.e vectors.

$$\text{similarity}(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

Where, A=book1(vector1)

B=book2(vector2)

A_i=rating given to the book 'A' by a user 'i'

B_i=rating given to the book 'B' by a user 'i'

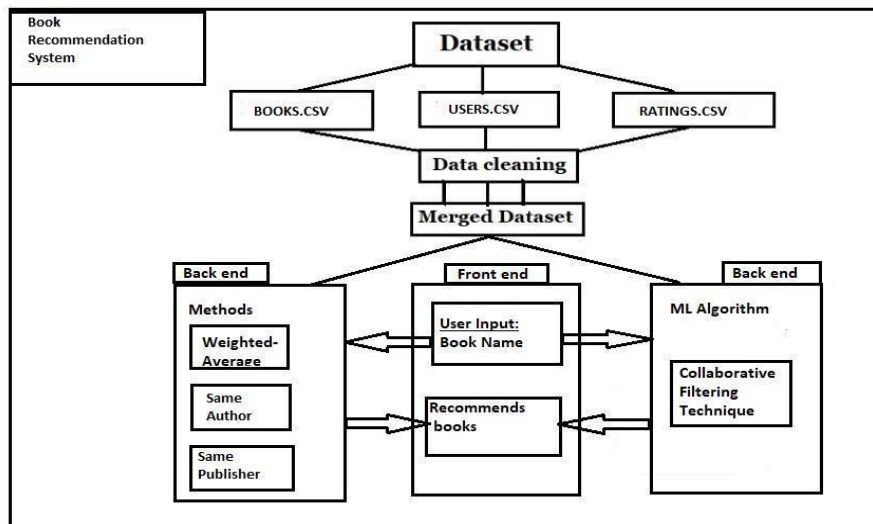


Fig.(1) : System Architecture for Book Recommendation System using Collaborative Filtering

5.1.2. Euclidean Distance Similarity(EDS):

Euclidean Distance Similarity(EDS) is based on the distance between two vectors. EDS measure compares absolute values and determines that the best variant of similar vectors is two absolute identical vectors with zero distance between them.

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (2)$$

5.2: Weighted-Average:

Weighted average[7][8] is a calculation that takes into account the varying degrees of importance of the numbers in a data set. In calculating a weighted average, each number in the data set is multiplied by a predetermined weight before the final calculation is made. A weighted average can be more accurate than a simple average in which all numbers in a data set are assigned an identical weight.

$$W = \frac{Rv + Cm}{v + m} \quad (3)$$

where, W=weighted-average score

R=average rating for the book

v=number of ratings for the book

C=the mean value of average-rating of all the books

m=minimum ratings required to get recommend

5.3: Same Author and Publisher:

As shown in the fig.3, users can search for the books of the same author or publisher by giving the name of the book as input. Author or Publisher of the input book is extracted from the dataset and top-n rated books of his/her are suggested to the users.

5.4: Web-application:

On the-other hand, we have developed a software web-application using flask, a python framework for backend and Html and Css for frontend and deployed in heroku (PaaS) cloud.

6.Results and Analysis:

Based on the information records of 278858 users, 242135 books, 1149780 ratings from the goodreadbooks kaggle datasets, unsupervised machine learning model i.e. collaborative filtering technique is trained and real-time web application is developed. The following takes the book title “Harry Potter and the Sorcerer’s Stone (Harry Potter (Paperback))” as an example input to show our proposed system recommended books as shown in the Fig.(2). and also to compare the performance of similarity measures.

6.1.Performance Comparison:

Figure. (2) illustrates the performance of two similarity measures CVS and EDS in terms of weighted-average of ratings with a varying number of recommendations. Same input book is given to both the measures and CVS is recommending the books with more rating than EDS and also as the size of the recommended books increases, weighted-average value for both similarity measures is decreasing. Finally, from Fig. 2, it is clear that the performance of collaborative filtering[CF] technique is better when CVS measure is used rather than EDS.

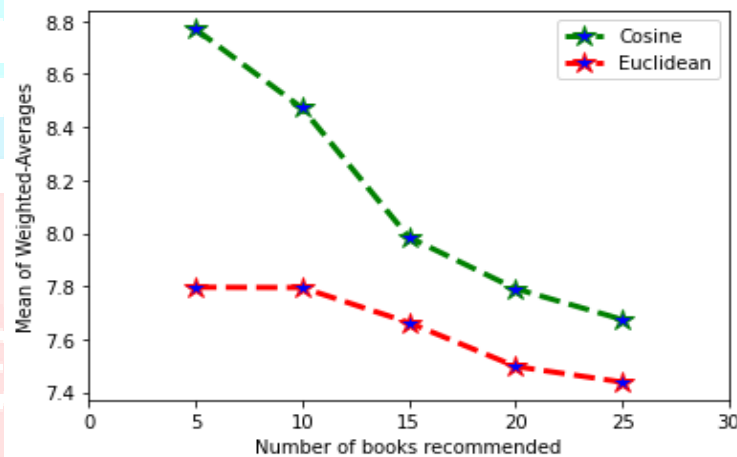


Fig.(2): Performance of two different similarity metrics

6.2. Evaluation Metric:

Evaluation metrics are used to evaluate learning algorithms[9] and form an important aspect of machine learning. In recommendation, we are predicting output, the error measures are fairly different here. As usual, the error metrics are obtained by comparing the predictions of the models with the real values of the target variables and calculating the average error.

6.2.1: Root Mean Square Error(RMSE):

To compute the RMSE, we first take the square of the difference between the actual and predicted values of every record[9]. We then take the average value of these squared errors. If the predicted value of the i th record is $y(i)$ and the actual value is $x(i)$, then the RMSE is:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x(i) - y(i))^2}{N}}$$

Where, N=number of datapoints

The merged dataset (Fig.1) is divided into 20% test data i.e. 206255 and 80% train data i.e. 825018, RMSE value is 7.9727.

As shown in Fig.(3), our proposed system takes the book title “Harry Potter and the Sorcerer's Stone (Harry Potter (Paperback))” as an example input to show the recommended books using CF technique as shown in Fig.(3).

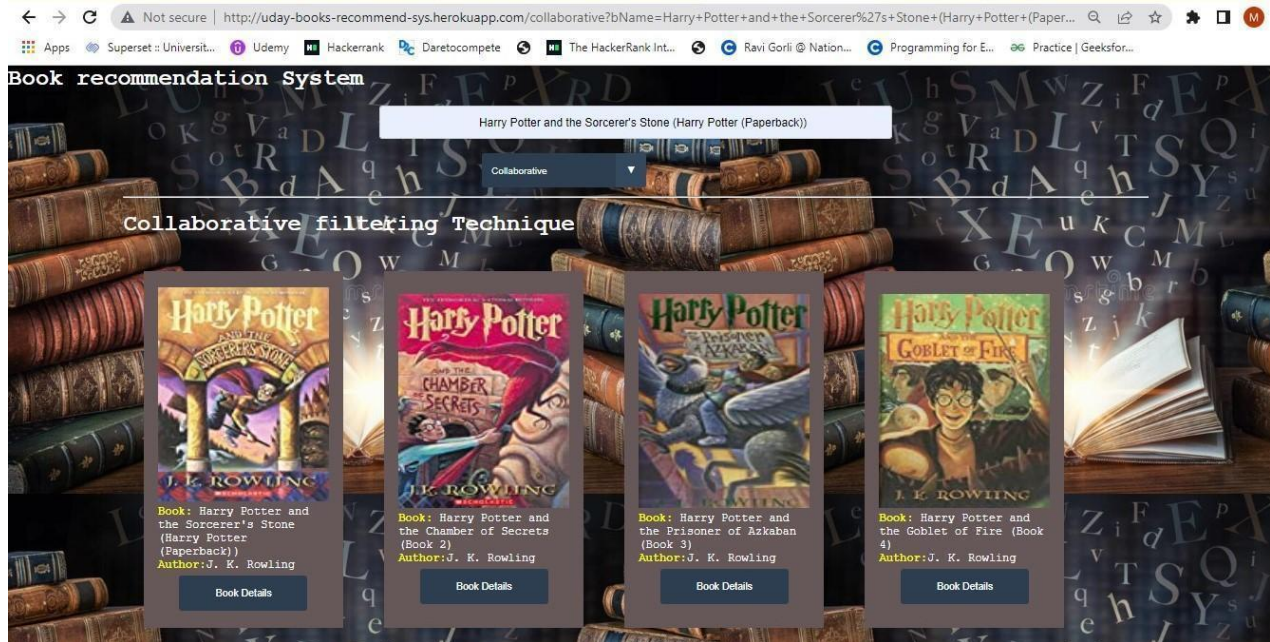


Fig.(3): Books Recommended using CF

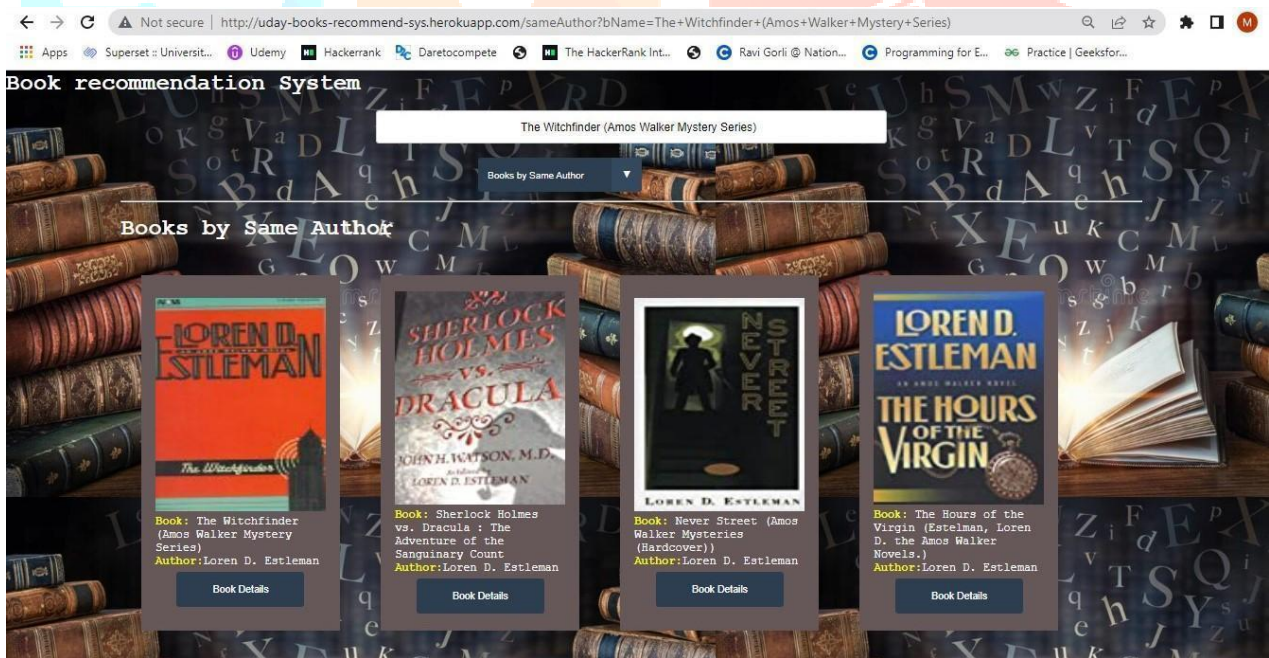


Fig.(3): Books by the Same Author

As shown in the Fig.(3), our proposed system takes “The Witchfinder (Amos Walker Mystery Series)” as an input and results in the n top-rated books by the author of the given input book. By clicking on the “Book Details” button as shown in Fig.3 results in the book name, Author, Publisher, Year-Of-Publication of the result book.

Conclusion:

As it is difficult for people to select the best books from a large set of books, this paper provides a modern solution i.e. a book recommender system based on item-based collaborative filtering [CF] technique which provides relevant recommendations. This paper also shows that using cosine similarity as a similarity measure gives best results when compared to euclidean distance, as euclidean distance gives the similarity value even if there exists null values. The CSV gives the best results when the null values are not present in the dataset. RMSE score is 7.9727.

In future, Collaborative Filtering can combine with content-based Filtering technique for more accurate recommendations. The proposed work can be applied for other domains like movies, music and some other.

References:

- [1]. Online Book Recommendation System using Collaborative Filtering (With Jaccard Similarity) To cite this article: Avi Rana and K. Deeba 2019 J. Phys.: Conf. Ser. 1362 012130 .(IOP)
- [2]. College Library Personalized Recommendation System Based on Hybrid Recommendation Algorithm, Yonghong Tian, Bing Zheng, Yanfang wang, Yue zhang, Qi Wu(2019) Procedia CIRP 83 (2019) 490-494(SCIENCEDIRECT)
- [3]. Online book recommendation system Conference Paper · September 2015 DOI: 10.1109/ICECCO.2015.7416895
- [4]. Asanov D. (2011) Algorithms and Methods in Recommender Systems. Berlin Institute of Technology Berlin, Germany
- [5]. Online book recommendation system by using collaborative filtering and association mining, Dr. Bharti Joshi (Ph.D) Scholar 10.1109/ICCIC.2015.7435717, 21 March 2016 (ICCIC)
- [6]. Machine Learning based Efficient Recommendation System for Book Selection using User based Collaborative Filtering Algorithm, 10.1109/ICISC47916.2020.9171222, 19 August 2020. (IEEE)
- [7]. Book Recommendation System through Content Based and Collaborative Filtering Method Ms. Praveena Matthew 1, Ms. Bincy Kuriakose 2, Mr. Vinayak Hegde 3, 10.1109/SAPIENCE.2016.7684166
- [8]. International Conference on Advanced Computing Technologies and Applications (ICACTA-2015) Introducing Hybrid Technique for Optimization of Book Recommender System Manisha Chandaka, Sheetal Giraseb, Debajyoti Mukhopadhyay, * a, b, c Department of IT, Maharashtra Institute of Technology, Kothrud, Pune 411038, India. (SCIENCEDIRECT)
- [9]. Aggarwal, C.C. (2016). Recommendation System: The Textbook. XXI, 29 p. ISBN 978-3-319-29657-9