Book Recommendation System

Disha Jethva, Aishwarya Rule, Dr. Kranti Ghag
Information Technology Shah and anchor Kutchhi Eng. Collage, Chembur, Mumbai-400080

ABSTRACT: The data mining technology is used to extract valuable information from various sources and make that information into understandable format. This paper proposed an interest-based book recommendation system. Recommendation system plays importantrole in e-commerce website to make good choices for users. In this paper we used three approaches for recommendation and that techniques are Collaborative filtering, Content-based and hybrid based technique. Collaborative filtering is used to measurethe performance of similarity measures in recommending books to user while content-based technique shows the highly purchased and top ten books using cosine similarity algorithm based on title and last the hybrid-based approach gave best similar book of user interest. In the end result shows that the classification report and best books as recommended for user.

I.   INTRODUCTION

Recommendation System(RS)plays an important role in raising the income of the e-commerce system. From last few years RS is used in many dominions like online shopping as well as entertainment. Buying something online or from e-commerce website is very crucial task for users, to overcome this problem RS is the good choice to implement [1]. Now we are in 21st century and it is a digital world and all the things is using internet services. RS is very favorable in the field of eLearning. ELearning is a field where every electronic devices (like laptops, smart phones, Desktop, etc.) are used [2]. There are many scenarios which happen in day to day life and in those scenarios we can use RS for eg. If the user doesn't have any about the item and wants to select some book for reading. Due to large number of sources user could not get appropriate item. So to overcome these problems, we have proposed system like recommendation system [3].
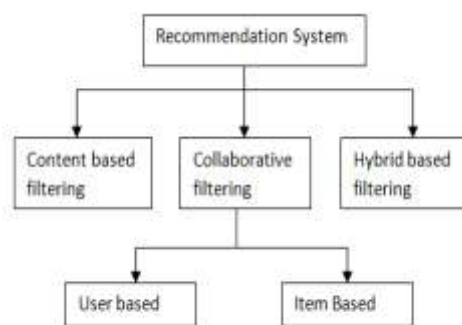

Fig. 1 Block diagram of Recommendation System

Various recommendation systems are proposedand applied in various domains and all domains are predicted correctly according to user interest. The recommendation can be made by various approaches namely collaborative filtering, content based filtering and hybrid based filtering [4].

The most common approach of the recommendation system is called collaborative filtering. Collaborative filtering filters the information and data collected by the system from the users [5]. Collaborative filtering method creates a model based on past experience of the user. This method works between the user and item [6]. The other approach of recommendation system which we have used was content based filtering. Content-based filtering, suggested items to the users on the basis of relation between user profiles and content item. And least but not the last approach was hybrid-based filtering.Here on the basis of the some parameters, most popular choices are found out. [7].

II.   LITERATURE REVIEW

As we know literature review is the primary area of the project and it is a mandatory part of the research project.

Book recommendation system using data cleaning and association rule mining use apriori algorithm. Apriori algorithm is a traditional

---

association mining algorithm. The procedure is composed of linking (matrix like operation) and pruning (drop the unnecessary intermediate results). Association rules has a concept called transaction [1].

A personalized recommendation model, it takes popularity and inverse popularity as one of the factors of interest. There is a negative correlation between inverse popularity and popularity. To elaborate briefly, inverse popularity is a scenario in life where people who are interested in unpopular things are more likely to be friends and have higher similarities. People who are interested in unpopular books or unpopular authors often have the same preferences [6].

A hybrid based recommendation algorithm based on context-awareness and social network. This algorithm firstly obtain the users current task and context information, and then calculate the context-entropy and context weight to represent the influence degree of different context factor on resource selection.Traditional person similarity is combined with the similarity calculated by neighborhood-based social network layer to get the hybrid user similarity. The final prediction results are obtained by the hybrid weighting process of the users' prediction score in different context [2].

Book recommendation for the eLearning usingcollaborative filtering. CF is used cosine similarities which are used to calculate the maximum number of similar documents amongst the available documents. This approach is used to access the degree of any two documents that can be compared based on their properties. This approach is widely applicable where a huge amount of information is available and we want to find valuable information from this available information [3].

III. DATASET

The dataset used for this is taken from Data mining cup where you can get the data about books, authors and their titles along with clicks [4]. Werecommend the books to users based on cosine similarity, traditional collaborative filtering and formula based approach. The recommendations of a book are being affected by many variables such as itemID, sessionID, clicks, basket, order and main topic. This system uses good reads data set. The data set consist of three tables:item.csv, transaction.csv and evaluation.csv

- Item.csv – This table consist of records about books like author, title, main topic, publisher and sub topics.

- Transaction.csv – Transaction table consist of data for a session and respective click values, whether it was added to basket or not, whether it was bought during the session or not.
- Evaluation.csv – This table consist of only item id as a column.

IV. METHODOLOGY

For recommendation system, there are several approaches or many algorithms that we can apply. In this paper we have used three approaches namely collaborative filtering, content-based filtering and hybrid based filtering.

1. Collaborative Filtering (CF): Collaborative filtering systems focus on relationship between users and items. Similarity of items is determined by the similarity of the ratings of those items by the users who have rated both the items[3]. The approach is to group people with similar taste and feedback from some can be used to recommend items to the group.

Fig. 2 Collaborative filtering

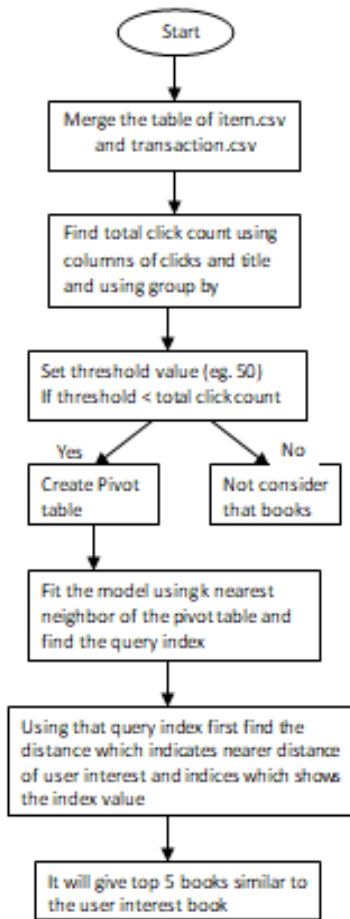In this project we used this approach and steps are as follows:

the items to bring out similar kind of items to the user. The system uses some information about items and also user's profile to see what user liked in the past. In this technique,we usedfollowing algorithms:
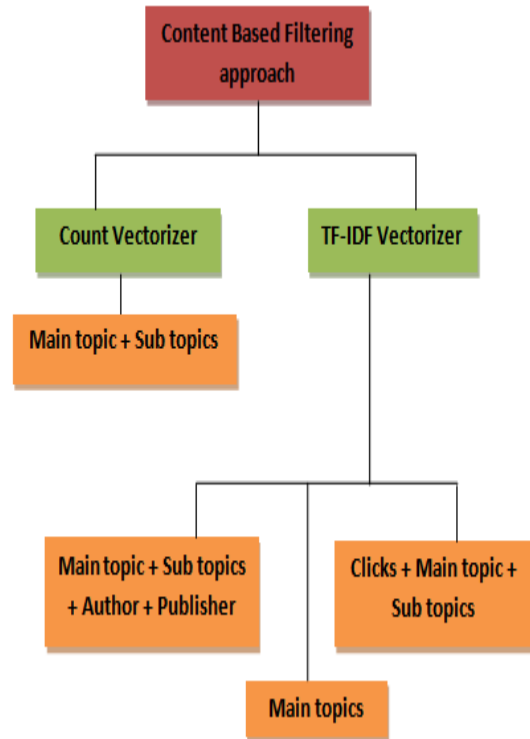

Fig. 4 Process of content-based filtering

Figure 4 shows the whole process of the content-based approach.

1.   TF-IDF Vectorizer:
    TF-IDF means Term Frequency - Inverse Document Frequency. Here Term Frequency indicates the frequency of a word in some document whereas the Inverse Document Frequency indicates frequency of a word in a set of documents (lower for higher frequency). The TF_IDF value is calculated for every word which increases with increase in frequency of every in document and gradually decreases with appearance of word in other documents.

2.   CountVectorizer:
    Machines can only understand numeric data. In this method, after manipulations like removing the stop words, transforming them to lower case, etc. we apply CountVectorizer converts text data to numerical data. It will only indicate about occurrences of a word.

Fig. 3 Process of collaborative recommendation

    In Figure 3, CF uses to find similarities between user and item. It is divided into two approaches namely Item-based and User-based. In user-based CF calculate similarities among all the users. In item-based CF calculates similarities between items that of users.

    First merge two tables of the dataset, after merging of tables find the click count using group by query using columns clicks and title. Once we got the total click count we have created pivot matrix table, it helps to show how many books are there in the basket in decimals and the nearest books to a particular book. After creating this pivot table, we used K Nearest Neighbor (KNN) for fitting the model and find the query index, using that query index we found distance between two items and index which shows top 5 books nearest book and their index values are returned.

2. Content-based Filtering (CBF): Content based recommendation system focuses on properties of

TF-IDF is better than Count Vectorizers because it not only focuses on the frequency of words present in the sentences but also provides the importance of the words.
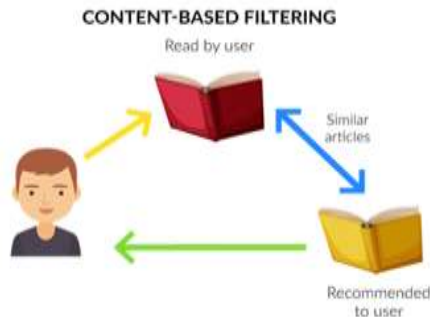

Fig. 5 Content-based filtering

3. Hybrid-based Filtering: It is a technique which has features of both collaborative as well as content-based. This technique overcomes the problem which are occurs in CF and CBF. We have used formula based approach in this technique.
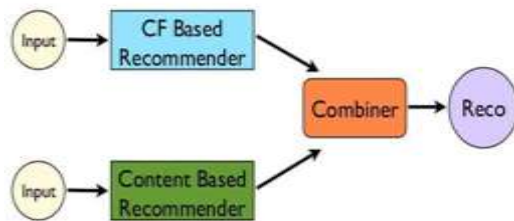

Fig. 6 Hybrid-based recommendation

Steps of the hybrid-based process are as follows:
Step 1:
Find the weighted average and for finding the weighted average we used the below formula:
Weighted average $= \dfrac{((R*v) + (C*m))}{(v + m)}$
Where, R is combine value of clicks, v is combine value of order, c is mean value of clicks and m is a quintiles value of order.
Step 2:
Sort the table according to weighted average and apply MinMaxScaler if the value of weighted average is high MinMaxScaler convert that value into between 0 and 1.
Step 3:
After that we created score column for that we used formula

Score = weighted average * 0.5
Step 4: According to score value it will recommend best book from the given data. The score value of the book is high that book is consider as a best book from the above formula.

V.   RESULT AND ANALYSIS

This section represents results for the techniques used for recommendation system experiments.Here we are presenting the accuracy measures of the proposed approach by constructing a confusion matrix obtained by testing on the test data using.

1. Collaborative filtering: The results of this approach is represented by the below comparison of the parameters.

K value	Accuracy	Precision	Recall	F1-score
1	0.91	0.95	0.95	0.95
2	0.95	0.95	1.00	0.98
3	0.94	0.95	0.99	0.97
4	0.95	0.95	1.00	0.98
5	0.95	0.95	1.00	0.98

Fig. 7 Comparison of parameters on the bases of k values

The top five books nearer to the distance of user interest book that we got from this approach is
Recommendation for ('Finde den Weg. Spannende Labyrinthe für Kinder ab 5', 'YNV', '[5AF,YBG,YBL,YNV]'):
1: ('Turtles All the Way Down', 'YFM', '[5AP,YFN,YXD,YXE,YXH,YXZ]'),with distance of 2.5287205440704952e-08:

2. Content-based filtering: The results of this approach are represented by:

Fig. 8 Recommended book from columns main topic

Fig. 9 Recommended book from columns main topic and subtopics

Fig. 10 Recommended book from columns main topic and subtopics and clicks

3. Hybrid-based filtering: Representation of the results using this technique is according to score best book is

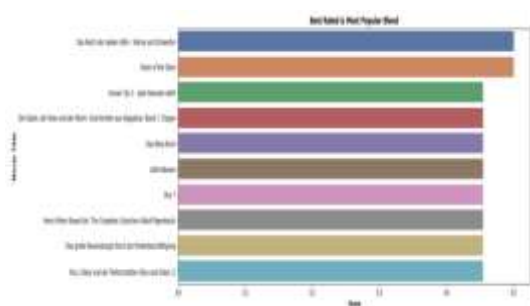| title	normalized_weight_average	score
448 54	Das Reich der sieben Höfe - Sterne und Schwerter	1.000000	0.500000

Fig. 11 graphical representation

OUTPUT

Fig. 12 Home Page suggesting popular choices

Fig. 13 Recommendation for "Mao und das Vermächtnis von Atlantis"

VI.   CONCLUSION

This paper mainly concentrated on the Book Recommendation for three popular approaches Collaborative Filtering Content-based and hybrid-based filtering. The main challenge faced is cold start that is the dataset available was too large so the filtering of the data was a laborious task. For collaborative filtering approach the data was reduced by considering a threshold value for the total click count to be 50. Coming to the content-based approach the algorithm used was TF-IDF Vectorizer and CountVectorizer and the entire dataset was scaled using fit transform. The result obtained was plotted using a sigmoid function with different combinations of columns. Also, the different combination of columns gives different results and compared different CSVs and found the duplicates. The third approach was hybrid based approach which used the features of both CF as well as CBF. The rank was obtained from the above mentioned formula.

VII.   FUTURE SCOPE

Future work is that the proposed recommendation system can be extended to cross domain with context-awareness to provide better recommendation accuracy and overcome coldstart and sparsity problems. In the future, we plan to explore this further and add more topic modeling

algorithms such as non-negative matrix factorization (NMF). Furthermore, we plan to include their preferences to provide recommendations that are more personal and satisfying.

REFERENCES

[1]. Zheng Wang, Shaohua Li, Jingying Feng and Yiduo Liang, "A Book Recommendation Algorithm Based on Data Cleaning and Association Rules," 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021.

[2]. H. Liu and N. Jiao, "A Hybrid Book Recommendation Algorithm Based on Context Awareness and Social Network," 2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), 2020.

[3]. T. Anwar, V. Uma, and Shahjad, "Book Recommendation for eLearning Using Collaborative Filtering and Sequential Pattern Mining," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 2020.

[4]. M. Kommineni, P. Alekhya, T. M. Vyshnavi, V. Aparna, K. Swetha, and V. Mounika, "Machine Learning based Efficient Recommendation System for Book Selection using User based Collaborative Filtering Algorithm," 2020 Fourth International Conference on Inventive Systems and Control (ICISC), 2020.

[5]. X. Hu, J. T. D. Ng, C. Yang, and S. K. W. Chu, "Personalized Book Recommendation to Young Readers," Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, 2020.

[6]. J. Li, T. W. Xu, and J. Zhou, "A personalized Recommendation Algorithm for College Books based on User Interest," 2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI), 2020

[7]. R. Albalawi, T. H. Yeap, and M. Benyoucef, "Toward A Real-Time Social Recommendation System," Proceedings of the 11th International Conference on Management of Digital EcoSystems, 2019.

[8]. N. R. Pinnapareddy, "Deep Learning based Recommendation Systems."

[9]. S. A. Stein, G. M. Weiss, Y. Chen, and D. D. Leeds, "A College Major Recommendation System," Fourteenth ACM Conference on Recommender Systems, 2020.

[10]. M. M. Y and M. K, "AI based Book Recommender System with Hybrid Approach," International Journal of Engineering Research and, vol. V9, no. 02, 2020.