BOOK RECOMMENDATION SYSTEM: A SYSTEMATIC REVIEW AND RESEARCH ISSUES

Article *in* Huazhong Keji Daxue Xuebao (Ziran Kexue Ban)/Journal of Huazhong University of Science and Technology (Natural Science Edition) · June 2021

3 authors, including:

Raghavendra C K
B. N. M. Institute of Technology
18 PUBLICATIONS **31** CITATIONS

BOOK RECOMMENDATION SYSTEM: A SYSTEMATIC REVIEW AND RESEARCH ISSUES

Pooja M Kerudi
UG Student
Computer Science and Engineering
BNM Institute of Technology
Bangalore-560070,India
Email -pooja17cs064@bnmit.in

Pratheeksha M
UG Student
Computer Science and Engineering
BNM Institute of Technology
Bangalore-560070,India
Email -pratheeksha17cs069@bnmit.in

T Sai Srujan
UG Student
Computer Science and Engineering
BNM Institute of Technology
Bangalore-560070, India
Email -sai17cs105@bnmit.in

Mr. Raghavendra C K,
Assistant Professor
Computer Science & Engineering,
BNM Institute of Technology,
Bengaluru, 560070, India
Email -raghavendrack@bnmit.in

***Abstract*— Recommendation systems are intelligent systems that employ a large database of information to offer the most accurate and appropriate products to consumers depending on their interests. Electronic- commerce (E-commerce) replicates the behavior of the human being and help to positively change the behavior of the sellers and consumers. Book Recommendation System (BRS) recommends a set of books to users based on their previous ratings. The current work focuses on content-based filtering, collaborative filtering, and hybrid filtering, along with many other ways for making recommendations. This study also highlights restrictions such as sparsity, cold start, and so on.**

Keywords*— *Book Recommendation system, Hybrid Filtering technique, Cold start, Sparsity, Content Based Filtering, Demographic Based Recommender system, Collaborative Filtering technique.

I. INTRODUCTION

Technology's use in everyday life has increased due to the impact of its fast advances. Technology is seen in almost every sector. Technology usage in banking, finance, media, education, etc. has boomed up in the last two decades. Likewise, online shopping i.e. Electronic Commerce (E-Commerce) has made shopping more convenient. E-commerce deals with the exchanging of various types of commodities and services through an electronic network. In the past decade, the usage of E-Commerce websites has increased substantially. The contribution of E-Commerce in total retail sales increased from 5.1% in 2007 to 16.0% in 2019.

With the increase in online shopping, the variety of items sold online has increased exponentially. Also, the number of people shopping online is rising day-by-day. Due to this, the users find it difficult to search the products of their interest. The tools such as Recommendation systems come into picture. Recommendation systems can be used to guide the users with items that they might prefer. They filter out suitable items from a large collection of items. The usage of Recommendation systems make it easier for the users to spot relevant items and improve their experience in online shopping. The online shopping websites can improve their revenue with the usage of Recommendation systems. These recommender system uses users interest, applies some filtering techniques and recommends the items based on the data. In some of the E-commerce websites, the system uses demographic data such as age, gender, location to recommend the items to the customer that he/she might prefer.

II. TYPES OF RECOMMENDATION SYSTEM

Recommendation system behaves like human experts which suggest the desired items to the users based on decision making ability. Nowadays, these Recommendation Systems (RS) are used in many sectors such as financial, medical, agricultural, educational, entertainment sectors etc. Recommendation Systems (RS) takes huge amount of data from the users, filters the data according to the algorithms specified by the user and then recommends the data based on given data. These Recommender systems can also be used for personal interest such as in YouTube, where each user gets a specific set of personalized recommendations based on the past behavior. A variety of recommendation system techniques work on various sources of data. The methodologies used to filter the data are Hybrid Filtering, Content Based Filtering Demographic based filtering and Collaborative Filtering .

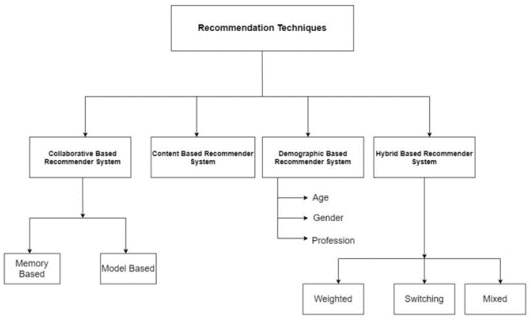

Figure 1: Types of Recommendation Techniques

A. Collaborative Filtering: Collaborative filtering is the popularly used filtering technique which is used to find the like-minded users. Based on the user's recent behavior, the system predicts the desired item for the user in this filtering. [1].This filtering evaluates user similarities based on their ratings and recommends new items based on inter-user comparisons.

Collaborative filtering techniques are further divided into two main categories:

i. Memory based filtering techniques: This is a type of collaborative filtering which approaches to solve the problem by considering the entire dataset This algorithm looks for people who are similar to the current user. The user to whom the recommendation is being made is referred to as active user. Users who have comparable behavior and transaction history to the active user are referred to as similar users [2]. To find the similarity between the users, the system uses a similarity measure such as Cosine similarity, Pearson similarity, Jaccard similarity, Correlation, etc. [3]. The similarity based on Correlation ranges in [-1, 1]. Correlation value 1 denotes the highest similarity and correlation value -1 denotes the least similarity between a pair of users. This technique is also called user-based collaborative filtering.
ii. Model based techniques: In order to build a system that is much faster and scalable the model-based technique can be used. In this technique "model" is the main keyword which refers to build the model based on the ratings present in the dataset. The main advantage of this technique is that the model doesn't use the complete dataset but predicts the scalable output. It is implemented using either explicit information such as ratings or implicit ratings such as user interactions and behavior. Modeling can be done using a variety of machine learning methods such as classification, cluster analysis, rule-based approaches, and so on [2].

B. Content Based Recommendation:
These recommender systems suggest items which are similar to the items bought by the user. In this technique, similarity scores between two items are calculated, based on which similar books are suggested to the active user. Any attribute or feature of the item can be considered while calculating similarity [4].
These recommender systems relate various items based on their features. User can use two methods in order to implement content-based similarity:
i. Cosine Similarity- It is a metrics used to find the similarity between the two items without depending upon the size of dataset. The two vectors have to be plotted on the multi-dimensional array, and measure the cosine angle between two vectors which can be helpful for the user to find the similarity between two items. Since, multi-dimensional space is used large-sized dataset can also support this technique [3].

ii. Decision tree classification- In this approach, the whole dataset is divided into many small numbers of sets which helps in predicting the output faster. The tree is constructed based on the past ratings which are rated by the present or previous user and based on the contents which are present in the dataset. Since, it uses the previous data of the users this technique produces the most appropriate output [5].

C. Demographic Based Recommender System: It is a kind of Recommender which divides the users into demographic classes using demographic attributes such as age, profession, gender, education etc. It works same like collaborative and content based filtering except it doesn't require any historical data. Various personal data taken during user registration can be used in this approach. For example, if the demographic attribute considered is profession of the users, then all researchers are recommended with a particular set of research related books, all the teachers are recommended with a particular set of academic books. One more parameter which can be used in this technique is the gender, based the gender specified by the user while registering the system tries to predict the book/items. Hence, the demographic based filtering helps the users to predict the efficient product without the usage of historical data [1].

D. Hybrid Recommendation System: To make more accurate recommendations, hybrid recommender systems integrate the features of two, three, or more filtering algorithms. These recommender systems integrate the good features of each technique to provide recommendations more effectively [6]. Hybrid filtering is further divided into
i. Weighted hybridization-In this hybridization technique, weights are allotted to each technique based on their performance. Initially, equal weights are allotted to each technique. The technique which proves to be more efficient than the other(s) techniques, gradually get greater weights than the other(s).
ii. Switching hybridization- Works based on the swapping of two or more recommendation technique, and uses the efficient technique to recommend the item.
iii. Mixed hybridization- This technique collects all the results from different recommendation technique at the same time and combines them to provide desired output. This technique is used when the data set is very large. In this type of hybridization each element of the hybrid should be able to produce desired output and then the system assigns some weights to the output. The output which has highest value is considered to recommend the predicted output.

III. LITERATURE SURVEY ON VARIOUS RECOMMENDATION SYSTEM

Recommendation system is a trending application in recent times. Recommendation systems are very important both from the user perspective and the company perspective. Users can conveniently shop using the applications with the recommendations being received from the recommendation systems. On the other hand, an application's usage, reachability and revenue can be improved tremendously with the usage of recommendation systems. Many experts have tried to improve upon the current Book Recommendation Systems. They have proposed various new, innovative ideas for the creation of efficient recommendation systems. This survey paper discusses some of such unique ideas.

Madhuri Kommineni, P.Alekhya, T.MohanaVyshnavi, V.Aparna, K Swetha, V Mounika [3] discussed that User Based Collaborative Filtering technique along with the cosine rule is more effective to predict the desired books to the user. In User based filtering the system finds the similar preferences of several users and recommends the next book which like-minded user may like to read. This system is very helpful for the administration purpose as it collects the feedback from all the users, report them and analyze the items and recommends most desired output. User profile as well as item profile is maintained to find the "User

Behavior" which is very effective in finding the desired output. In order to build collaborative filtering recommendation they have used Singular Value Decomposition (SVD) model which helps to predict more efficiently and effectively. The quick sort algorithm is used to sort the dataset based on the keywords provided by the users after registering. Historical data should be maintained properly in this system.

Jayanti Rathnavel and Kavita Kelkar [1] proposed a personalized recommender for recommending books to the users. In this experiment, they combined the two popularly, extensively used recommendation techniques i.e. collaborative and content based techniques to build a hybrid recommender. They personalized the system by trying to understand the interests of the users such as favorite author, favorite genre, etc. They have addressed the overspecialization problem. Overspecialization is a limitation in which the recommended books are similar to those that the current user has already read. Using lightfm model, overspecialization is overcome, due to which the recommended set of books also contains the type of books not explored by the active user. It gives the opportunity to the active user to explore new kinds of books. The recommeder can learn the new interests of the active user.

Anand Shanker Tewari and Kumari Priyanka [7] in their paper proposed a The Book Recommendation System Based on Collaborative Filtering and Association Rule Mining (ARM) for College Students employs the User Based Collaborative Filtering technique to forecast the top n-rated books for students and academics. This system aims to help the students to find books based on the price ranges and publisher's name. The system employs categorization approaches, collaborative filtering based on user input, and association rule mining. Classification techniques are used to extract a set of rules and patterns in the data and classify the data to predefined classes, each class is processed independently while recommending. Similar people are detected using Pearson's similarity algorithm in user-based collaborative filtering. ARM determines the correlation of each users in the given dataset and associate the relation between users and finds the best suited items. ARM can also be used to discover interesting associations and relationships in the data, which can be used for user behaviour analysis. Based on these techniques, the system recommends books to the readers.

Praveena Mathew, Bincy Kuriakose and Vinayak Hegde [4] According to the author, combining content-based filtering with collaborative filtering produces more effective and efficient results. Along with these two techniques, associative rule is used to predict the desired items from a large collection of items. This method aims to tackle the problem of sparsity by combining the techniques of Content Based Filtering, Collaborative Filtering, and Associative Rule Mining. The system also implements keyword based recommendation in which, the users enter keywords related to their interests and the system compares these words in the datasets to recommend the books. Equivalence class Clustering and bottom up Lattice Traversal are discussed in this paper (ECLAT), which aims to find frequently read sets of books in an efficient way. ECLAT performs using Depth First Search (DFS), thus scanning the dataset only once and consuming less time compared to other algorithms.

Kitti Puritat and Kannikar Intawong [8] have proposed a model for book recommendation system that uses Support Vector Machine (SVM). They took into account a variety of factors, including title similarity and book bibliographic information like author, year, category, number of books, etc. This model was specifically designed for usage in small libraries. SVM is a supervised machine learning model that can be used to solve classification and regression applications. The SVM is trained using three sources of data i.e. title similarity, Dewey Decimal Classification (DDC) for classification and bibliographic features. The model was found to perform considerably well.

Huayong Liu and Nianlai Jiao [9] have proposed a hybrid recommendation system with the usage of context awareness and social network. Various contextual factors that affect the user choice on books are obtained through the context aware layer such as gender of the reader, time of borrowing the book, etc. A user-book-context matrix is established to represent the contextual theme suitable for book recommendation. The contextual factors are associated with the book type in the matrix and then context aware computing is performed to obtain entropy and the weight of each contextual factor. On the other hand, user-to-user similarity is calculated based on Pearson similarity, based on which nearest k users are considered. The books to which the active user has not rated are scored using the corresponding scores of these k nearest users. The obtained scores are combined with context weights to obtain final scores, upon which recommendations are made. It is opined in the paper that with the usage multiple other context factors and multi-dimensional context factors, the system can be further improved.

JiabeiLi, TianweiXu, Juxiang Zhou [10] demonstrates how to use the hybridization method to effectively use content-based filtering and collaborative filtering techniques. The combination of popularity, inverse popularity with similarity and duration of borrowing of book is considered to measure the user's interest and likeability on those books. Inverse popularity highlights that the users who like unpopular books have similar interest. While applying for inverse popularity Borrowing Time has been calculated. Borrowing Time refers to the time between the borrowal and return of the book. It reflects the reader's interest on the book. If a person's borrowal duration on a book is short, he/she might be more interested on it. On the other hand, if it is longer, he/she might be less interested on it.is the time. Including all the above mentioned parameters, the recommendation is done. The scalability issues are overcome in this system with the usage of cloud.

Dharna Patel, Harish Patidar [6] have proposed a Recommendation Solution for Online Book Portal and have explained the need of cloud computing while recommending the books. The value-added feature in this paper is that the system gets the profession of the user while registering into the system. To recommend the book collaborative filtering technique and content-based filtering techniques are being utilized. Cloud computing has been used in this paper as

dataset is very large and it is not always possible to store it in local disk, which is very difficult to recover when in case of any loss. In order to secure data, one can adopt cloud computing, which is also known as a storing centre because it maintains enormous datasets. This system of recommendation is more suitable to the readers who require the best book for general purpose rather than specific purpose. Raghavendra et al, [18,19] provided an study of existing techniques, similarity metrics and research opportunities in this area.

IV. TYPICAL DATASETS FOR BOOK RECOMMENDATION SYSTEM

Sl no.	Dataset	Link
1.	Goodreads –books	https://www.kaggle.com/jealousleopard/goodreadsbooks
2.	Amazon Top 50 Bestselling Books 2009 – 2019	https://www.kaggle.com/sootersaalu/amazon-top-50-bestselling-books-2009-2019
3.	Top 270 Computer Science / Programing Books	https://www.kaggle.com/thomaskonstantin/top-270-rated-computer-science-programing-books
4.	Highly Rated Children Books And Stories	ttps://www.kaggle.com/thomaskonstantin/highly-rated-children-books-and-stories
5.	7k Books	https://www.kaggle.com/dylanjcastillo/7k-books-with-metadata
6.	GoodReads 100k books	https://www.kaggle.com/mdhamani/goodreads-books-100k
7.	Google Books Dataset	https://www.kaggle.com/bilalyussef/google-books-dataset
8.	Books Dataset	https://www.kaggle.com/saurabhbagchi/books-dataset
9.	Book-Crossing: User review ratings	https://www.kaggle.com/ruchi798/bookcrossing-dataset
10.	Goodreads books/author data	https://www.kaggle.com/brosen255/goodreads-books

Table 1: Description of dataset along with their login credentials

V. RESEARCH ISSUES

A. Cold start – It is one of the most serious problems that Recommendation Systems encounter. When a person joins the system for the first time, he or she will have no history data. Thus, it is difficult to recommend books to him/her. This issue is called cold start problem [1,4]. The same problem arises when there is a new book in the system, where it is difficult to recommend that book to the existing users. Many researchers have tried to solve cold start problem in various approaches. Using the demographic data of the users, this issue can be overcome [1]. Rather than using historical data, users' demographic data might be utilized to recommend books to them. This problem can also be solved with the help of a Knowledge Graph Convolutional Network. [11]. Many other solutions to this problem is necessary to provide significant results.

B. Sparsity – This issue arises when there is a lack of ratings and reviews by the users. With lesser ratings and reviews it is difficult to understand the user's taste and give recommendations. Thus, reducing the effectiveness of the recommendation system [1,4,7,10]. A possible solution is to use Knowledge Graphs to alleviate this problem [12]. Another solution would be to use matrix factorization. This issue opens up new opportunities for improvement.

C. Trust issues – This issue arises when certain users have lesser history and it is difficult to recommend books to them. Also, it occurs when it is difficult to decide the amount of weightage to be given to reviews and ratings. The users might have varying tastes, thus making it difficult to recommend books to them. That is, we do not know how much we can trust the existing ratings of users with respect to books. Social network data can be used to reduce the impact of this issue [9]. There is a scope of research in this issue.

D. Scalability -With the increase in the number of books and users in the system, the scalability issue arises. The system requires more and more resources for recommendation [8]. Also, the performance of the system might not be significant with the increase in number of books and users. Thus, it is necessary to develop recommendation models that can face the scaling up of data.

VI. CONCLUSION

With the adoption of shopping and reading online, the recommendation systems have become a necessity. There are multiple types of recommendation approaches that can be used to build a recommender. These types of recommenders can be combined together in the hybrid approach to get better results. Also, there are many issues faced by that arise with respect to recommendation systems such as sparsity, cold start, scalability, etc. Thus, more extensive research and development of models in this field is necessary to alleviate these issues.

REFERENCES

[1] Jayanti Rathnavel, Kavita Kelkar. "Personalized Book Recommendation System". In International Journal Of Engineering And Computer Science ISSN: 2319-7242, Volume 6 Issue 4 April 2017, Page No. 21149-21153.

[2] P. H. Aditya, I. Budi, Q. Munajat "A Comparative Analysis of Memory-based and Model-based Collaborative Filtering on the Implementation of Recommender System for Ecommerce in Indonesia : A Case Study PT X"

[3] MadhuriKomminen,P.Alekhya,T.MohanaVyshnavi,V.Aparna,KSwetha,VMounika. "Machine Learning based Efficient Recommendation System for Book Selection using User based Collaborative Filtering Algorithm". International Journal of Advanced Research in Computer and Communication Engineering, 2018..

[4] Ms.PraveenaMathew, Ms. BincyKuriakose, Mr.VinayakHegde."Book Recommendation System through Content Based and Collaborative Filtering Method". In Department of Computer Science And Engineering, international journal of innovations in engineering and technology (IJIET), ISSN:2319-1058,vol.2, issue 2, pp1-7, April 2017.

[5] Snehalata B. Shirude.Et.al worked on "Classification of Library Resources in Recommender System Using Machine Learning Techniques: 52nd Annual Convention of the Computer Society of India, CSI 2017, Kolkata, India"(January 2018)

[6] Dharna Patel, Dr. Harish Patidar. "Hybrid Recommendation Solution for Online Book Portal". In International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887 Volume 6 Issue V, May 2018.

[7] AnandShanker Tewari, Kumari Priyanka. "Book Recommendation System Based on Collaborative Filtering and Association Rule

Mining for College Students". In 2014 International Conference on Contemporary Computing and Informatics (IC3I).

[8] Kitti Puritat, Kannikar Intawong. "Development of an Open Source Automated Library System with Book Recommendation System for Small Libraries". In 2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON).

[9] Huayong Liu, Nianlai Jiao. "A Hybrid Book Recommendation Algorithm Based on Context awareness and Social Network". In 2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE).

[10] JiabeiLi, TianweiXu, Juxiang Zhou. "A personalized Recommendation Algorithm for College Books based on User Interest". In 2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI).

[11] MadhuriKomminen,P.Alekhya,T.MohanaVyshnavi,V.Aparna,KSwetha,VMounika. "Machine Learning based Efficient Recommendation System for Book Selection using User based Collaborative Filtering Algorithm". International Journal of Advanced Research in Computer and Communication Engineering, 2018.

[12] Anoop A, N Ayush Ubale. "Cloud Based Collaborative Filtering Algorithm For Library Book Recommendation System". Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020).

[13] E. Uko Okon, B. O. Eke, P. O. Asagba. "An Improved Online Book Recommender System using Collaborative Filtering Algorithm". In International Journal of Computer Applications (0975 – 8887) Volume 179 – No.46, June 2018.

[14] Dharna Patel, Dr. Harish Patidar. "Hybrid Recommendation Solution for Online Book Portal". In International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887 Volume 6 Issue V, May 2018.

[15] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, Minyi Guo1."Ripple Net: Propagating User Preferences on the Knowledge Graph for Recommender Systems1". In arXiv: 1803.03467v4 [cs.IR] 25 Aug 2018.

[16] Cheng Zhao, Chenliang Li, Rong Xiao, Hongbo Deng, Aixin Sun. "CATN: Cross-Domain Recommendation for Cold-Start Users via Aspect Transfer Network". In arXiv: 2005.10549v2 [cs.IR] 23 May 2020.

[17] HUAXUAN ZHAO, HONGCHEN WU, JINGZHI LI, HUAXIANG ZHANG AND XINJUN WANG. "Dtree2vec: A High-Accuracy and Dynamic Scheme for Real-Time Book Recommendation by Serialized Chapters and Local Fine-Grained Partitioning". SPECIAL SECTION ON ADVANCED DATA MINING METHODS FOR SOCIAL COMPUTING, January 2020.

[18] Raghavendra C K, Srikantaiah K.C, Venugopal K. R, " Personalized Recommendation Systems (PRES): A Comprehensive Study and Research Issues", International Journal of Modern Education and Computer Science(IJMECS), Vol.10, No.10, pp. 11-21, 2018. DOI: 10.5815/ijmecs.2018.10.02

[19] R. C K and K. C. Srikantaiah, "Similarity Based Collaborative Filtering Model for Movie Recommendation Systems," *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2021, pp. 1143-1147, doi: 10.1109/ICICCS51141.2021.9432354.