

Book Recommendation System using Machine Learning Algorithm

²MANOHAR, ^{1*}.Md. IRSHAD HUSSIN B

²Master of Computer Applicatios,University B.D.T College of Engineerig

Davangree, Karnataka, India

¹PROFESSOR, Department of Master of Computer Applicatios,University B.D.T College of Engineerig Davangree,

Karnataka, India

ABSTRACT:

Recommender Systems are intelligent systems which are used as an expert in making decisions in real life problems. They have replicated the human experts and positively affected the e-commerce by changing the behavior of customers and sellers. Book Recommender Systems (BRS) help the librarians in the management of library catalog efficiently. It supports the readers in choosing the best book for them. Merchants implement the BRS to manage their inventory and gain more profit. In this paper, we have discussed traditional techniques of recommendation, machine learning techniques and their categories i.e. supervised, unsupervised, semisupervised and reinforcement learning. Also, Machine Learning (ML) techniques used for the book recommendation and their effect on book recommender systems have been discussed. The work will help the researchers in exploring new dimension for recommendation technology in general and book recommendation in particular.

which he may like or which are related to his area of study if he is a student or a teacher. Similarly, a book recommender applied in e-commerce domain suggest the merchants about the different books and help him to manage his inventory. It also helps the user in purchasing the most appropriate book for him considering various criteria like his preferences, cost and other features of the book.

Machine learning (ML) is one of the fascinating and fastest growing field of this era. ML techniques are being implemented in almost every domain of computing. They have been implemented in BRS to improve the predictive and recommendation accuracy of the system. These techniques have been utilized for extraction and classification of book features. Human sentiments and their choices have also been extracted using these. After finding sentiments, choices, browsing history, and features of books similarity is calculated and recommendations are generated. This paper is a study on machine learning techniques which have been implemented on book recommenders. In this paper we have reviewed how ML techniques have been implemented in BRS, what is the benefit of using a particular technique, what other techniques can be used for further improving the recommendation systems and what are the applications of BRS.

I. INTRODUCTION

The growth of the internet and advancement in computing technologies has touched every aspect of human being. Nowadays computers are being implemented in almost every domain of life to replicate the manual system and to reduce the workload of humans. Recommenders are a type of intelligent systems which are implemented to replicate human experts. They have been implemented in various real-life applications to assist people in decision making. In e-commerce recommenders have been implemented to help the buyers in choosing the most suitable product for them and to increase the profit of merchants by selling more products. Book recommenders have been designed considering the changing pace of time. They are implemented in libraries for their maximum utilization. By mining the borrowing records of libraries it suggests the librarians the books to be purchased. On the basis of browsing history of a library user a recommender suggests him the books

II. LITERATURE SURVEY :

Ms. Praveena Mathew, Ms.bincy kuriakose and Mr. Vinayak hedge [1] proposed a Book Recommendation System (BRS) through the combined features of content based filtering (CBF), collaborative filtering (CF) and association rule mining to produce efficient and effective recommendation. The existing systems lead to extraction of irrelevant information and lead to lack of user satisfaction. So, they proposing a hybrid algorithm, which combines two or more algorithms, to help the recommendation system to recommend the book based on the buyer's interest. They use association rule mining algorithm, ECLAT (Equivalence class clustering and bottom up lattice traversal). ECLAT will helps to find out the frequent item set. It uses depth first searching technique. In one scan, it will categorise. Cosine similarity is used for the similarity measuring in content and collaborative filtering. They use item-

item filtering in collaborative filtering. The basic finding that achieved through this proposed work is to recommend the books based on the buyer's interest and increase the productivity and credibility. Using association rule mining algorithm to finds interesting association and relationship among large data set of books and provides an efficient recommendation for the book.

Yongen Liang and Shiming Wan [2] proposed a method, which can mine products by understanding the user's preferences. It is a personalised technology with collaborative filtering. It is book recommendation system, which is for a university library. Here only provide the recommendation service to the registered users. The collaborative filtering uses both user-user filtering and item-item filtering. The important job of the collaborative filtering is to calculate the similarity of the books and users or reader then, recommend. Cosine similarity is using for the similarity measuring in collaborative filtering. Then find out or predict the rating for the particular book, which the targeted user may like or give. One of the most important problems of collaborative filtering is cold start. That is, when a new user joins then they have no data about that user. They have no previous purchase history or borrow history. Therefore, here they propose a solution that Expert and new book recommendation. Expert and new book recommendation module will recommend the books as if Best-selling, newbooks arrived, classical books... in short, it will recommend the books at the top rating or popular books.

Dharmendra Pathak, Sandeep Matharia and C. N. S. Murthy [3] proposed an efficient and best unique hybrid recommendation algorithm, by providing the recommendation more satisfying the user's desire. Here the hybrid recommendation is a combination of collaborative, content and context based recommendation algorithms. The main input of collaborative filtering is rating i.e, votes of so many people, content based data that is the information about the users like their interest, date of birth, priorities... and the context based data that is the behavioural datas like date, taste, mood, weather... Cosine similarity is using for the similarity measuring. There are subject priorities according to the user's previous history. If they purchase a book then check, the purchased book is different subject priority from the subject priority has already set? If yes, then reset the subject priority3 and then subject priority 2. The subject priority1 will not change. Based on calculations and results they concluded that the proposed Hybrid book recommendation algorithm is best among the others.

Ahmed.M. Omran [4] proposed a Hybrid Recommendation system that will answer for the questions like, which book to buy? Which financial service to choose? Which website to visit next? First phase, collaborative filtering that is based on user behavior by calculating the statistical correlation between

the internet users' profiles using pearson correlation Factor by considering the number of visits to varies websites for each user to estimate the type and the strength of correlation among users. Then, Second phase applies content based filtering according to the content of websites by computing the relative similarity between each pair of websites and build, a pairwise comparison matrix to find the most nearby websites to the most visited users' websites. In collaborative filtering, from the browsing history, collect the websites, which that user visited. Then make the user profile with this data and record how many times that particular user visited in each site. Also make neighbourhood that is, find the similar users to that particular user. Spearman statistical method is the way of finding the users that have a common behavior. Content based filtering is the second phase. Here by using the text data mining technique that commonly used in content-based technique i.e, TF-representation filter the data to predict items to users determine the similarity between websites by counting the words of the main pages and applying one of the data mining techniques to find the category to which website belongs. There are five criteria to set the similarity of each couple of websites i.e, Category, Service, Language, Rating, and Interactive. The Euclidean distance is using for similarity measuring.

Adli Ihsan Hariadi, and Dade Nurjanah [5] proposed a hybrid-based method that combines attribute based and user personality based methods for book recommender system. In this paper, they are implementing the MSVMSL (Most Similar Visited Material to the Most Similar Learner) method, and they are saying that, it is the best method among hybrid attributes based methods. The personality factor is used to find the similarity between users when creating neighbourhood relationships. The hybrid attribute will calculates the recommendation scores of rated books from neighbors using the similarity scores between a target book and its neighbors and between the active user and that user's neighbours. The score of book b from user u, denoted as score_b. This is for finding the Most Similar Visited Material to the most Similar Learner. It uses the values from both content and collaborative. Then use the result of hybrid as recommendation. That is the Most Similar Visited Material to the most Similar Learner.

Anand Shanker Tewari, Abhay Kumar and Asim Gopal Barman [6] proposed a book recommendation system based on combined features of content filtering, collaborative filtering and association rule mining. When a buyer search for a book, then it will be store as a purchase history or a search history. When the buyer is offline the recommendation perform some filtration for recommending to buyer and the results are stored in the buyers web profile. When the buyer comes online next time, the recommendations will be generated automatically. In content based filtering, web Usage Mining (WUM) is used to provide relevant information to the buyers.

web Usage Mining (WUM) typically extracts knowledge by analyzing historical data such as web server access logs, browser caches, or proxy logs. It helps to possible to model user behavior and, therefore, to forecast their future movements. web Usage Mining stores the user's behavior on the internet and processes that data. Item based collaborative recommendation Algorithm is using and Cosine similarity is using for the similarity measuring. Intersect the results from the association rule mining and the content, collaborative filtering.

Binge Cui and Xin Chen [7] proposed a novel book recommendation system. The readers will be redirect to the recommendation pages when they cannot find the required book through the library bibliographic retrieval system. It is an online book recommendation system for a library and it is based on web service. After login, a user search for a book with keywords like a book title, or with author name... at that time bibliographic retrieval system will search for books with the same keywords. If found any result in the recommendation system, then send these keywords to web Books Retrieval Module. In web Books Retrieval Module, it search on the online bookstore based with keywords by creating accounts on these online book stores like amazon... by the librarian or the admin of the online book recommendation system. Therefore, when the keyword comes to the web retrieval module, it searches as user on these online bookstores by login in. The result getting from these online bookstores will give to the user as recommendation results. According to the recommendation from users, the statistic and analysis module will calculate the value of that particular book. Then according to this value of book, the Auto-Order Module will produce a book order automatically based on the analysis results. When the purchased books have shelved, Book Storage System will send a report to the Short Message and Email Notification Module. Then it will notify the readers that have recommended the purchased books using Message and Email server.

Kumari Priyanka, Anand Shanker Tewari and Asim Gopal Barman [8] Personalised Book Recommendation System Based On Opinion Mining Technique. An online book recommendation; especially consider the specific features of the book that a particular user already purchased. Here, not only consider the feature but also consider the reviews given by the user for the books. So, here uses the technique that opinion mining or sentiment analysis to classify the reviews or comments from the different users for different book into positive or negative. For this, naïve bayes algorithm will perform the text classification. The classification of the review will helps to identify the user's preference and the books rating.

Raymond J. Mooney and Lorie Roy [9] proposed a Content-Based Book Recommending Using ML tools for

Text Categorization. They describe a content-based book recommending system that utilizes information extraction and machinelearning algorithm for text categorization. Learning individualized profiles from descriptions of examples, on the other hand, allows a system to uniquely characterize each patron without having to match his or her interests to another's, Items are recommended based on information about the item itself rather than on the preferences of other users. They have been exploring content-based book recommending by applying automated text-categorization methods to semi-structured text extracted from the web. The current prototype system, LIBRA (Learning Intelligent Book Recommending Agent), uses a database of book information. The system then learns a profile of the user using a Bayesian learning algorithm and produces a ranked list of the most recommended additional titles from the system's catalog. Overall, the results are quite encouraging even when the system gives relatively small training sets. LIBRA is an initial content-based book recommender, which uses a simple

III. METHODOLOGY

The proposed system in Fig. 1 used a clustering technique to develop the recommender system. Fig. 1 shows three parts named data acquisition, preprocessing, and clustering techniques. The datasets were collected from the Goodreadsbooks repository of kaggle in this research. Though Goodreads-books repository of kaggle contains seven datasets, only four datasets (Books.csv, Book_tags.csv, Ratings.csv, and Max_Rating.csv) were considered for this experiment. The preprocessing technique was applied after merging all datasets where we removed the lower-rated books and developed a new dataset for analysis. Finally, the clustering technique was applied for recommending books to those users who stay in proximity to a specific cluster. Besides, a user can then search for a book through a query interface, and results in listing recommended books (Fig. 6).

A. Data Acquisition

The dataset was collected from the GoodReads book dataset repository. It has 10,000 rated data of popular books. This data set consists of 7 tables named Books.csv, Geners.csv, Book_tags.csv, Max_rating.csv, Ratings.csv, to_read.csv, and Tags.csv, where we used Books.csv and Book_tags.csv as book dataset and Ratings.csv and Max_Rating.csv as user rating dataset. The description of the datasets are as follows:

- Books.csv- it has attributes like an author, book_isbn number, rating and contains 10K books.
- Book_tags.csv- it has 596K rows and attributes are goodreader_book_id and tag_id.

- Ratings.csv- it has attributes like user_id,book_id, and rating and contains about 9,00,000 rows.
- Max_Rating.csv- it has similar attributes as Rating.csv. But the number of rows is about 500K.

B. Data Preprocessing

Unstructured noisy text in the data is needed to be preprocessed to make them analyzable. To do the analysis, the dataset needs to be cleaned, standardized, and noise-free. Fig. 2 shows that most of the books were rated 4 or above. We want to recommend only top-rated books. So we remove all the rows having a rating less than 4. It shows us that 68.89% of books were rated 4 and above. Thus our cleaned dataset becomes compact, standardized, and noise-free.

C. Clustering Techniques

K-mean algorithm is used as a cluster partition algorithm where each partition is considered as a k cluster. It is an agile algorithm applied in cluster assessment, feature discovery, and vector quantization. In this experiment, the k-mean algorithm begins with selecting the numbers of k cluster of books. Each book is assigned to the nearest cluster center and moved from the cluster center to cluster average and repeated until the algorithm reaches to convergence state.

Fig. 3 shows the cosine similarity function which calculates the cosine of the angle between two non-zero vectors (vectors A and B).When these vectors align in the same direction then they produce a similarity measurement of 1. If these vectors align perpendicularly then the similarity is 0, whereas two vectors align in the opposite direction will produce a similarity measurement of -1.



Fig.2. the pie chart of rating Dataset

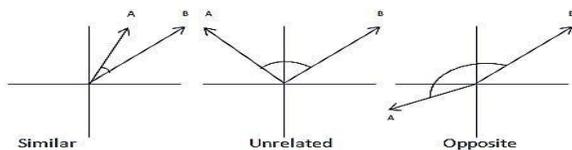


Fig. 3. Cosine Similarity and Cosine Distance Functions.

IV. COMPARATIVE STUDY OF DIFFERENT ALGORITHMS

| YEAR | AUTHOR | PURPOSE | METHOD MENTIONED |
|------|---|--|--|
| 2016 | Ms. Praveena Mathew, Ms.bincy kuriakose and Mr. Vinayak hedge | Book Recommendation System Through Content Based And Collaborative Filtering Method | Combined Features Of Content Based Filtering (CBF), Collaborative Filtering (CF) And Association Rule Mining |
| 2018 | Yongen Liang and Shiming Wang | The Design And Implementation Of Books Recommendation System | Collaborative filtering. |
| 2009 | Binge Cui and Xin Chen | An Online Book Recommendation System Based On Web Service | Content based |
| 2018 | Ahmed.M. Omran | A Novel Recommender System For Websites | Hybrid method. |
| 2017 | Adli Ihsan Hariadi, and Dade Nurjanah | Hybrid Attribute And Personality Based System For Book Recommender Recommendation | Hybrid-based method |
| 2014 | Anand Shanker Tewari, Abhay Kumar and Asim Gopal Barman | Book Recommendation System Based On Combine Features Of Content Based Filtering, Collaborative Filtering And Association Rule Mining | Combined Features Of Content Filtering, Collaborative Filtering And Association Rule Mining. |

| | | | |
|------|------------------------|---|---------------|
| 2009 | Binge Cui and Xin Chen | An Online Book Recommendation System Based On Web Service | Content based |
|------|------------------------|---|---------------|

Table 1: comparative study of different algorithms

V. EXPERIMENT RESULTS :

As Shani and Gunawardana say [14], experiments on recommendation systems are divided into three categories: a) offline experiments: when user behavior is simulated i.e. the process of system making predictions and recommendations and the user corrects them or uses these recommendation is simulated, b) user studies: analyze the quantitative and qualitative measurements and observe the behavior of a group of subjects that will interact with the system, c) online evaluation: the system is used by real users on real tasks and then analyzed. In our research we decided to use an online survey because of to the best of our knowledge there is no available database suitable for our task and because online evaluation can not be applied yet. Therefore we decided to analyze the performance of our system by creating an online survey.

We selected a group of independent readers and asked them to give score (0-10) on such parameters of the system like quality of recommendations and convenience and ease of use of the system. The higher score indicates the relevance of recommendations. Also additional field for user notes was added to the questionnaire.

Experiment results are demonstrated in Fig. 6. They show the average of user opinions. It can be seen that users liked the easy of use of the system i.e. the speed of getting recommendations (89%) while the quality of recommendations was estimated at 77%.

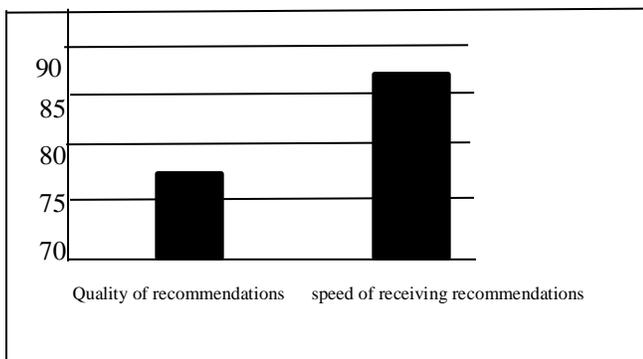


Fig. 6. Experiment results

Although the promising results of online survey users indicated several problems related to system. One of them was that sometimes the same book appears in different categories which makes recommendations less useful. This is because some of the genres are very closely related to each other and therefore the items of that genres can appear in both of genres.

VI. CONCLUSION :

A Recommendation System is an act of trying to regulate the future work of the E-commerce industry. The prediction of user’s interest of books or any other item could yield remarkable profit and change in the field of decision-making, retail business, hotel business, tourism, digital content, movie databases. From this literature survey, we found that, the most suitable algorithm for predict books for users by considering their preference and by avoiding cold start problem is, the item based collaborative filtering with opinion mining on reviews. This algorithms will be a great benefit for users as well as the owner.

VII. FUTURE WORK:

In this paper, we have discussed the book recommender system and its impact on library management, benefits of BRS to librarians, students, other library users, and booksellers. We also have discussed traditional recommendation algorithms, machine learning techniques and their major classes. We have also discussed MLT applied in BRS and their impacts on BRS. In future work, we will design a framework for applying MLT in BRS, and will study in which stage of recommendation each MLT is used, how other MLT’s which have not been used in BRS can be incorporated to improve BRS. What are the challenges in implementing MLT’s in BRS and what are the advantages and disadvantages of using MLT’s in BRS.

REFERENCES :

[1.] Book Recommendation System Through Content Based And Collaborative Filtering Method by, Ms. Praveena Mathew, Ms.Bincy Kuriakose & Mr. Vinayak Hedge.

- [2.] The Design And Implementation Of Books Recommendation System By, Yongen Liang And Shimmingwan [3.] Nova :Hybrid Book Recommendation Engine By, Dharmendra Pathak, Sandeep Matharia And C. N. S. Murthy [4.] A Novel Recommender System for websites by, Ahmed.M.Omran.
- [5.] Hybrid Attribute and Personality Based Recommender System for Book Recommendation By, Adli Ihsan Hariadi, Dade Nurjanah.
- [6.] Book Recommendation System Based On Combine Features Of Content Based Filtering, Collaborative Filtering And Association Rule Mining By, Anand Shanker Tewari, Abhay Kumar And Asim Gopal Barman.
- [7.] An Online Book Recommendation System Based on web Service By, Binge Cui, Xin Chen.
- [8.] Personalised Book Recommendation System Based On Opinion Mining Technique By, Kumari Priyanka Anand Shanker Tewari Asim Gopal Barman.
- [9.] Content-Based Book Recommending Using Learning For Text Categorization By, Raymond J. Mooney And Lorie Roy
- [10.] Developing a hybrid collaborative filtering recommendation system with opinion mining on purchase review By, Idong Yun, Danial Hooshyar, Jaechoon Jo and Heuseok Lim
- [11.] Collaborative filtering-based recommendation system for big data By, Jian Shen, Tianqi Zhou and Lina Chen
- [12.] An Optimized Item-Based Collaborative Filtering Recommendation Algorithm By, Jinbo Zhang, Zhiqing Lin, Bo Xiao, And Chuang Zhang
- [13.] Systematic Approach for Cold Start Issues in Recommendations System By, M. Sarumathi, S. Singarani, S. Thameemaa, V.Umayal, S.Archana, Ms.K.Indira and Dr.M.K.Kavitha Devi
- [14.] G. Shani and A. Gunawardana, "Evaluating Recommendation Systems," Springer, 2011.
- [15] G. Adomavicius and A. Tuzhilin, "Towards the Next Generation of Recommender Systems: A Survey of the State-of- the-Art and Possible Extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005.
- [16] X. Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," *Adv. Artif. Intell.*, 2009.
- [17] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Comput.*, vol. 7, no. 1, pp. 76–80, 2003.
- [18] S. Kumar Jain and C. Rana, "Building a Book Recommender system using time based content filtering," *WSEAS Trans. Comput.*, vol. 11, no. 2, pp. 27–33, 2012.
- [19]"Whichbook.net." [Online]. Available: <http://whichbook.net>. [Accessed: 20-May-2015].
- [20] "Lazylibrary.com." [Online]. Available: <http://www.lazylibrary.com>. [Accessed: 24-May-2015].
- [21] M. S. Pera, Y. Ng, and N. Conde, "Personalized Book Recommendations Created by Using Social Media Data," in *Web Information Systems Engineering - WISE 2010 Workshops*, 2010, pp. 1–14.
- [22] Guy, N. Zwerdling, I. Ronen, D. Carmel, and E. Uziel, "Social Media Recommendation Based on People and Tags," in *Proc. of SIGIR*, 2010, pp. 194–201.