Volume: 14 Issue: 03 Pages: 5486-5492(2022) ISSN: 0975-0290

Enhancing Personalized Book Recommender System

Abdulgafar Usman
Department of Computer Science, Usmanu Danfodiyo University, Sokoto, NIGERIA.
Email: abdulgafarusman80@gmail.com
Abubakar Roko
Department of Computer Science, Usmanu Danfodiyo University, Sokoto, NIGERIA
Email: abroko@yahoo.com
Aminu B. Muhammad
Department of Computer Science, Usmanu Danfodiyo University, Sokoto, NIGERIA
Email: aminu.muhd@yahoo.com
Abba Almu
Department of Computer Science, Usmanu Danfodiyo University, Sokoto, NIGERIA
Email: almu.abba@udusok.edu.ng

ABSTRACT

Recommender systems (Rs) are widely used to provide recommendations for a collection of items or products that may be of interest to user or a group of users. Because of its superior performance, Content-Based Filtering (CBF) is one of the approaches that are commonly utilized in real-world Rs using Time-Frequency and Inverse Document Frequency (TF-IDF) to calculate document similarities. However, it computes document similarity directly in the word-count space. We propose a user-based collaborative filtering (UBCF) method to solve the problem of limited in content analysis which leads to a low prediction rate for large vocabulary. In this study, we present an algorithm that utilises Euclidean distance similarity function, to solve the identified problem. The performance of the proposed scheme was evaluated against the benchmark scheme using different performance metrics. The proposed scheme was implemented and an experimentally tested by using the benchmark datasets (Amazon review datasets). The results revealed that, the proposed scheme achieved better performance than the existing recommender system in terms of Root Mean Square Error (RMSE) which reduces the errors by 29% and also increase the Precision and Recall by 51.4%, and 55.8% respectively in the 1 million datasets.

Keywords - **Recommender System, Content-Based, Collaborative Filtering, Personalized Recommendations, Similarity Function.**

Date of Submission: Jun 30, 2022 Date of Acceptance: Sep 21, 2022

1. INTRODUCTION

The rapid growth of the World Wide Web increases the difficulty of finding information related to users' information needs. Since the 1990s, efforts have been towards tackling this problem through the use of recommender systems.

Recommender systems are information filtering systems that deal with the problem of information overload [1]. By selecting vital information fragments from a large amount of dynamically generated information based on the user's preferences, interests, or observed behavior concerning the item [2]. It can also predict whether a specific user would prefer an item or not based on the user's profile. Furthermore, it benefits both service providers and users [3]. Commercial applications of recommender systems include e-commerce e.g., Amazon, eBay [4], fashion e.g., food and restaurants [5], social events and art e.g. movies, music, books [6]. In these areas, recommender systems provide excellent personalized suggestions that significantly increase the likelihood that a customer purchases a product or selects an item compared with non-personalized suggestions.

Recommender systems are categorized into four approaches such as Content-based, Collaborative, and Hybrid [7]. The Content-based filtering technique is based on the description of the items and a profile of the user's preferences [20]. It uses user-item similarities for recommendation purposes. In this kind of recommender system, the recommended items to a given user are the ones similar to those that were previously liked by the user [8]; [9]; [10]. While collaborative filtering (CF) is the most widely used system because it is used to recommend different types of products e.g. music, news, books, movies, etc. [11]. The basic concept behind CF is based on collecting and analyzing a large amount of information on user's preferences and predict what users will like based on similarity to other users. While hybrid combines two or more recommendation strategies in different ways to benefit from their complementary advantages.

Various studies have been proposed on book recommender systems to prevent low dimensional prediction rates, limited content analysis, new user problems, and overspecialization. For example, the study [12], proposed personalized book recommender system for bookstore management using a Time-Frequency and Inverse Document Frequency (TF-IDF) model for

Volume: 14 Issue: 03 Pages: 5486-5492(2022) ISSN: 0975-0290

calculating the similarity between books. However, the method sometimes recommends books that don't match what the user needs because the TF-IDF method only uses book titles to calculate the similarity between books.

In this work, an Enhance Personalized Book Recommender System (EPBRS) is presented for bookstore management to improve the quality of recommendations and prediction accuracy. The system incorporated Euclidean distance similarity function to the existing personalized book recommender systems with the use of EPBRS algorithm to recommend books that matches what the user needs The experimental study conducted indicated that compared to the existing PBRS system, the proposed EPBRS system improves the precision, recall as well as RMSE of the 1 million dataset used which in turn returns provide good interested books recommendations to the target users.

2. RELATED WORKS

[13] Investigated a book-recommendation system that uses a content-based approach to recommend products to a specific user by employed temporary dimension. The temporal dimension takes into account the number of times an item is liked by the user over some time and stores the counter for each item with an update whenever a user checks the items in their favorite links. The results show that the proposed book recommendation engine architecture provides users with a diverse and temporarily updated recommendation that is more useful and relevant. However, the system incorporates contextual information, takes multiple ratings, and provides a more flexible recommendation that could also extend into deferential domains.

[12] Proposed a personalized book recommender system for bookstore management using TF-IDF. The TF-IDF method is based on the bag-of-word (BoWs) model, where the titles of the books are converted to BoWs. The system sometimes recommends books that are not what the user needs because the TF-IDF method uses only the title to compute the similarity directly in the word-count space, which lead to low prediction rate. This makes the less relevant to the users.

[14] Suggested a novel approach to recommend the university member's customized book lists by using the k-means clustering technique to create clusters of device users as well as using user transaction history and then recommending custom book lists to users within those clusters. The algorithm used a questionnaire survey to evaluate an individual's booklist accuracy. The result shows the possibility of using the history of circulation activity to predict an individual member's current interest and build the personalized booklist that suits their interests. The suggested books are however a poor representation of user needs due to the lack of consideration of the actual book material.

[15] Proposed a library book recommendation system based on the loading of user-profiles and applied the association rules for model creation. It uses library loan records and the technique of association rule mining to recommend books in the digital library. The scheme uses an association rule mining system to make inferences and derives interesting rules as computer students are more likely to be interested in computer and math books rather than geography. Using the "Amazon analysis dataset," the algorithm was tested using precision and recall metrics. The result showed that the new association rule algorithm is suitable for use in a library to recommend books. But the method is restricted to finding interesting rules that suit the needs of users who don't frequently visit libraries and conduct machine transactions.

[16] Proposed a novel collaborative filtering recommendation algorithm based on user correlation and evolutionary clustering. Firstly, the score matrix is pre-processed with normalization, and dimension reduction is considered to obtain denser score data. Based on these processed data, clustering principles are generated and dynamic evolutionary clustering is implemented. Secondly, it considered nearest neighbors by applying user correlation to choose nearest neighbors to predict rating. The proposed method is evaluated using the Movie-lens dataset. Diversity experimental results demonstrate that the proposed method has outstanding performance in predicted accuracy and recommended precision. However, the system did not address the data sparsity.

[17] Proposed a neural collaborative filtering recommender method that integrates user and item auxiliary information. It used Auto Encoder to extract user features, and Gated Recurrent Unit with auxiliary information to extract items' latent vectors, respectively. Also, the attention mechanism is used to learn key information when extracting text features. Research suggests that the experimental results show that the GANCF model can get better results on the two data sets than others. By considering auxiliary information and applying it to deeper learning can improve the recommendation performance of the model. But, sometimes the interests of users change with time.

[18] Proposed an improvement to the Ant Collaborative Filtering Algorithm by introducing the K-means algorithm to capture an initial clustering phase of the users according to their preferences represented by pheromones, of which there are far fewer compared to the number of users. The experiments were conducted on several larger datasets, including Movielens 10M, Douban book, and NetEase music datasets. The results demonstrate its excellence over the ACF algorithm in both scenarios, i.e., the rating-based recommendation and the ranking-based recommendation. However, a common problem with K-means is determining the parameter K or the number of clusters.

3. MATERIALS AND METHODS

This section describes the proposed enhanced book recommender system.

3.1 Problem Formulation

The review of the related works was conducted which led to the identification of the research gap in the existing literature. The personalized book recommender algorithm for bookstore management based on the Term-Frequency and inverse document frequency (TF-IDF) method uses

Volume: 14 Issue: 03 Pages: 5486-5492(2022) ISSN: 0975-0290

title of the books in form of BoWs model to recommend books to users. But, the system sometimes recommends books that are not what the user needs because it only utilizes book titles to compute document similarity directly in the word-count space, which lead to a low prediction rate for large vocabulary.

3.2 Enhanced Personalized Book Recommender System (EPBRS)

EPBRS is a Personalized Book Recommender System that uses different approaches, including collaborative, hybrid, content-based, knowledge-based and utility-based filtering. In this paper, we incorporate collaborative filtering along with Euclidean distance. The EPBRs system processing can be mainly divided into three steps: Collecting user rating data matrix, selecting similar neighbors by measuring the rating similarity, and predicting target user ratings. The user rating data consists of users, items, and user ratings on observed items in the form m × n matrix, where m presents the total number of users and n presents the total number of items. Rm,n is the ratings given to item n by user m.

3.2.1 User-Book Rating Matrix

Table1 shows the rating matrix consisting of users rating on books. To find the m user that exhibited similar interest with other users. It can be seen that user1 (U1) and user3 (U3) prefer book1 over user2 (U2). In addition, when it comes to book3, both U1 and U3 dislike it more than U2. Also U1 is more similar to U3. Looking at book1, it can be seen that U3 and U1 both like it, while U2 does not. For book 3, both U1 and U3 disliked it, whereas U2 liked it as seen in Table 1. It is obvious that U1 is more similar to U3, so we get other books that U1 has read but U3 has not. As a result, book5 can be recommended to U3 because U1 has already read it five-time.

Table 1: User Books Ratings

Review ID	U1	U2	U3
Book1	5	2	4
Book2	3	4	3
Book3	1	4	1
Book4	2		?
Book5	5	2	?

3.2.2 Similarity Calculation

The similarity calculation is a fundamental stage in the recommendation process. In this case, the similarity of user × user are computed from table 1 using equation (1) and the results are presented as follows:

$$Sim\,(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \qquad (3.1)$$

Where, (*p, q*) refers to dimensional vectors between two users or items and *n* is the number of items.

$$Sim\ (P_1, P_1) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$$
$$Sim\ (P_1, P_1) = \sqrt{(5-5)^2 + (2-2)^2 + (4-4)^2}$$
$$Sim\ (P_1, P_1) = \sqrt{(0)^2 + (0)^2 + (0)^2}$$
$$Sim\ (P_1, P_1) = \sqrt{0+0+0}$$
$$Sim\ (P_1, P_1) = \sqrt{0}$$
$$Sim\ (P_1, P_1) = 0$$

Table 2: User × User Similarity Computations

User_j / User_i	A	B	C	D	E
A	0.000000	0.000000	0.000000	0.160768	0.240253
B	0.000000	1.000000	0.261204	0.366025	0.500000
C	0.000000	0.261204	1.000000	0.000000	0.158945
D	0.160768	0.366025	0.000000	1.000000	0.151342
E	0.240253	0.500000	0.158945	0.151342	1.000000

Consider Table 2 below we intuitively decided not to take all neighbors into account (neighbor selection) for the calculation of the predictions, we included only those that had a positive correlation with the active user (and of cause, had rated the item for which we are looking for a prediction) if we included all users in the neighborhood, this would not only negatively influence the performance concerning the required calculation time, but it would also affect the accuracy of the recommendation, as the ratings of other users who are not comparable would be taken into account.

3.2.3 Rating Prediction

After computing the similarity between user x and with any other users, select the k users with highest similarity values as set N. Then, we try to estimate the rating of item i by the user x and make sure that the set N consist only of users actually rated item i. finally, we make a prediction for user x and item i using a weighted average rating from the neighborhood as in equation (2):

$$r_{xi} = \frac{\sum_{y \in N} S_{xy} \times r_{yi}}{\sum_{y \in N} S_{xy}} \qquad (3.2)$$

Where S_{xy} is the similarity of user x and y.
r_{yi} is the rating of user y on item i
N is the set of k users most similar to x who has also rated item i
r_{xi} is the prediction for user x and item i

Volume: 14 Issue: 03 Pages: 5486-5492(2022) ISSN: 0975-0290

Some sample prediction computations results are demonstrated in Table 3 by using the equation (2) above and the EPBRS algorithm.

Where S_{xy} is the similarity of user x and y.

r_{yi} is the rating of user y on item i
N is the set of k users most similar to x who has also rated item i
r_{xi} is the prediction for user x and item i

Algorithm: EPBRS
Input: Dataset - -Amazon review dataset (Reviews, Metadata, and Items) containing user-item rating information's
Output: Predicted items' ratings

1. Load dataset
2. Perform visualization and exploratory analysis on the dataset
3. Feature engineering //remove features/columns and ratings that are not needed) for each user:
4. Randomly split the user's ratings into test/train //this method of train test split is to prevent data bias
5. Perform error analysis: for each item:
6. if item rating > 5 then,
   set item_rating = 5 for each user:
7. if num of rating of each item rated > 1: then,
   item_rating = mean of user's ratings of that item
8. Create a User by Item matrix from the review data.
9. Compute items seen/rated by each user
10. Create User × User review by items matrix:
11. Compute a user by user similarity matrix using Euclidean distance in equation (1)
12. Generate user ratings for all items using user by user matrix obtained in step 11 by using equation (2)

Lines 1-4 of the algorithm represent the data collection and cleaning of the Amazon book review data. The data would be represented as an m x n matrix.
Lines 6 - 7 of the algorithm is to perform around check if any item ratings are greater than five to ignore it and set the item rating to be equal to five for each user. Also if the number of each item rated is greater than one then, we take the average ratings of that particular item to minimize error analysis from the dataset.
Lines 8 – 9 of the algorithm is to create a user x item matrix from the review data and we compute items seen/rated by each user.
Lines 10 – 12 of the algorithm is to create a user x user matrix with a pivot table and in each column, we compute user x user similarity matrix using Euclidean distance to see the distance between the user ratings.
Finally, we generate user ratings for all items using the user matrix obtain in step 11.

4. EXPERIMENTS AND RESULTS

This section discusses the experimental setup to evaluate our proposed EPBRS system. The effectiveness of the system was evaluated by measuring the recommendation quality using precision, recall, and RMSE evaluation metrics.

4.1 Experimental setup

We used computer Intel (R) Core i5-4310 2.6-GHz with 8 GB memory running Windows 10 Pro. The proposed system was implemented in Python 3.9 with Pandas 0.8.0 Library, SCIKITLearn.

4.2 Data set

The Amazon review dataset will also be considered in evaluating the performance. This dataset consists of 12,886,488 book reviews associated with information about books, users, and their ratings. The time and effectiveness of each review are given and it can be easily extended to include more information about books, but not about users [19]. Moreover, the Amazon API delivers item lookup not limited to books and similarity lookup that returns a list of similar products (usually presented to customers as recommendations).

Amazon dataset was used in the two experiments. We compare our proposed approach with a recommendation based on bestselling books by [12] using the Amazon dataset. We used four fold cross-validation. We divided the users into the ratio of 3:7 where 70% of the existing users were used for training, while 30% of the users were used for test data. We 30% of the users as cold users in which interest had already been known.

Volume: 14 Issue: 03 Pages: 5486-5492(2022) ISSN: 0975-0290

Table 3: Prediction using weighted average ratings from the neighborhood

	Rating	Reviewer ID	Asin	Unix Review Time	Predicted Ratings
214051 82	3.0	A108M62RB1HTC0	1682307913	1486425600	4.759705
206353 53	5.0	A3RF010MX29BQP	1939713293	1435622400	4.590341
151845 90	5.0	A277TO3PKKNYDH	1479327573	1351382400	4.465197
220801 26	5.0	ALY4MQYVAEE4U	1494563258	1408924800	4.719279
152202 83	5.0	A3526B1LCK47X9	1480114480	1353456000	4.282797
667423 9	5.0	A3KAKFHY9DAC8A	0446549150	1421107200	4.262266
706098 8	5.0	A1WH1YZ5VGTKXG	0451475682	1423008000	4.486042
117578 95	5.0	A5FDYZB6MG2TT	0989104400	1375833600	4.302070
218456 48	5.0	A2L7N2U5Z316ZE	B000X1MX7E	1268438400	4.590768
175066 15	4.0	ASZ8NR9HIZGJF	1519028520	1475971200	4.315153

Table 4: Overall Result of our proposed scheme

Data Size	Model	RMSE	Average Precision	Average Recall
1 Million	PBRS	0.973	0.37	0.35
	EPBRS	0.683	0.53	0.53

{
Category:"Books New, Used & Rental Textbooks
Medicine & Health Sciences"
"Books Arts & Photography Music"
"Books Arts & Photography Music"
Tittle: Biology Gods Living Creation Third
Edition 10 (A Beka Book Science Series)
Mksap 16 Audio Companion: Medical Knowledge
Self-Assessment Program
Flex! Discography of North American Punk,
Hardcore, and Powerpop 1975-1985 A-M
Heavenly Highway Hymns: Shaped-Note Hymnal
Georgina Goodman Nelson Womens Size 8.5 Purple
Regular Suede Platforms Shoes
Principles of Analgesic Use in the Treatment of Acute Pain and Cancer Pain
(APS, Principles of Analgesic Use in the Treatment of Acute Pain and Cancer Pain)
MKSAP 15 Audio Companion"
Rank:"1349781, 1702625, 6291012, 2384057, 11735726, 2906939, 2236549".
Price: "92878, 000047715X, 4545,13765,116,555010, 477141"
Rating: "5, 5, 5, 5, 5, 5, 5, 5, 5, 5"
Asin: "92878, 000047715X, 4545, 13765, 116, 555010, 477141"
}

Fig. 1 Sample amazons book review

4.3 Result and Discussion

In this section, the results of the improved bestselling approach were presented and discussed.

4.3.1 Results

In the section below, results of the similar user's interest and bestselling book were presented. It was obtained that the result of similar user interest and best bookselling books was better than that of bestselling books approach alone. Table: 6 below presents an overall result of RMSE, Precision, and Recall respectively. Prediction ratings, the total number of ratings, and the actual number of ratings were obtained from the Amazon dataset.

The table 6: bellow shows the overall result of our proposed schemes.

4.3.2 Discussion

Across all the evaluations in Table 4, results show that the user interest similarity PBRS technique and the bestselling books approach to achieve better RMSE, Precision, and Recall in 1 Million datasets than the work [12]. In the 1million dataset, there is an improvement in RMSE, Precision, and Recall of 29.8%, 51.4%, and 55.8% as presented in Fig. 1, Fig. 2 and Fig. 3 respectively. It can be seen from the results that, the EPBRS model achieves the lowest RMSE values than PBRS with higher average precision and recall for the 1 million dataset. Therefore, it shows better prediction performance which helps in recommending books of interest to the users.

The reason for this performance is because the PBRS moderately penalized high rating items since they can be of interest to the active users. It ignores the penalty on the items with fewer ratings to improve their chances of being predicted. It also indicated that taking into account the dataset's item popularity level played a better role in predicting items of interest to users. As a result, the proposed not only enhances the accuracy of rating prediction but can also help to predict more interesting bestselling books to users.

Volume: 14 Issue: 03 Pages: 5486-5492(2022) ISSN: 0975-0290

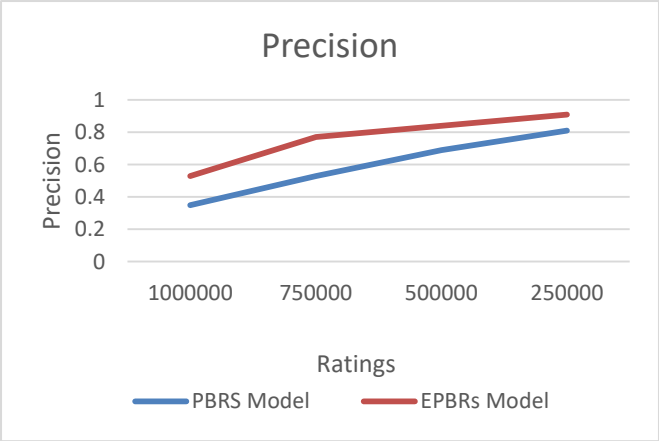

Fig. 2 Precision 1 Million dataset size

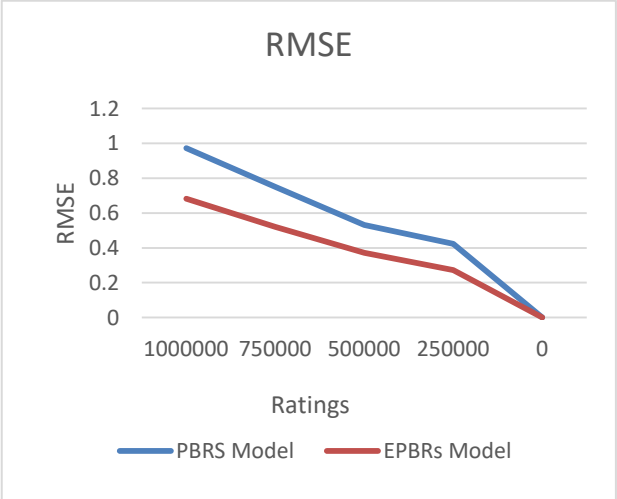

Fig. 3 RMSE 1 Million dataset size

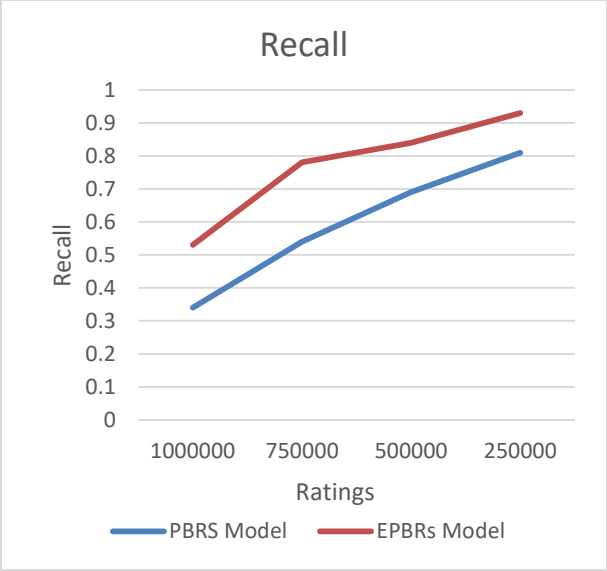

Fig. 4 Recall 1 Million dataset size

5. CONCLUSION

We proposed an Enhanced Personalized Book Recommender System (EPBRS) by utilising the Euclidean Distance Similarity Function. The scheme employs Euclidean distances similarity function to locate users with similar interests and books of interest. A user is recommended one of their books of interest that are bestselling books to users. The book's ratings were used as feature sets for bestselling predictions. The evaluation was performed using the "Amazon review dataset" and the results showed that the RMSE, Precision, and Recall improved by 29.8%, 51.4%, and 55.8% respectively in the 1 Million datasets. However, in our future work, we will incorporate different machine learning and clustering algorithms to allow regrouping of all books based on the ratings and user preferences and then study the comparative results.

REFERENCES

[1] M. Prem, and S. Vikas. *Recommender Systems.* (Encyclopedia of Machine Learning), 2010.

[2] C. Pan, and W. Li. Research paper recommendation with topic analysis. In computer design and Application *IEEE* 2010, 4-264.

[3] P. Pu, L. Chen, and R, Hu. A user-centric evaluation framework for recommender systems. *In Proceedings of the fifth ACM conference on Recommender Systems* (RecSys' 11), ACM, New York, NY, USA; 2010, P. 57-164.

[4] S. Sanjeevan, S. Alireza, R. Hossein, and M. Asad. Recommender systems in e-commerce. *In Proceedings of the World Automation Congress (WAC). IEEE*, 2014, 179– 184

[5] H. Kang, and J. Seong, SVM and collaborative filtering-based prediction of user preference for digital fashion recommendation systems. *IEICE Transactions on Information and Systems,* 2007, 2100 –2103.

[6] L. George, and C. Petros. A hybrid approach for movie recommendation. Multimedia Tools and Applications. *Conference on Innovative Applications of Artificial Intelligence* 2008, 55–70.

[7] J. Bobadilla, F. Ortega, A. Hernando, and A. GutiéRrez. Recommender systems survey. *Knowledge-Based System*, 9(46): 2013, 109 - 132.

[8] H. Zamani, and A. Shakery. A language model-based framework for multi-publisher content-based recommender systems. *For*

Volume: 14 Issue: 03 Pages: 5486-5492(2022) ISSN: 0975-0290

springer International Journal on Information Retrieval, 2(1): 2018, 369-409.

[9] U. Hanani, B. Shapira, and P. Shoval. Information filtering: Overview of issues, research, and systems. *User Modeling and User-Adapted Interaction,* 11(3), 2001, 203–259.

[10] P. Lops, M. Gemmis, and G. Semeraro. Content-based recommender systems: State of the art and trends. *In information Retrieval.*2011, 73–105, Springer.

[11] S. Gong and J. Softw. A Collaborative Filtering Recommendation Algorithm Based On User Clustering and Item Clustering. *Conference on Innovative Applications of Artificial Intelligence.* 2010, 745-752.

[12] J. Sarun and K. Paween. Automatic non-personalized book recommender algorithm for Bookstore shelf management: *The 3rd International Conference on Digital Arts, Media, and Technology (ICDAMT IEEE)*: 2018, 1-5.

[13] R. Chhavirana and K. J. Sanjay. Building a Book Recommender system using time-based Content Filtering. *WSEAS Transactions on Computers Issue* 2(11): 2012, 49-78.

[14] M. Suthathip and M. Songrit. A recommendation model for personalized book lists. *In Communications and Information Technologies (ISCIT) International Symposium on IEEE,* 2010, 389-394.

[15] P. Jomsri. Book recommendation system for digital library based on user profiles by using association rule. *In Innovative Computing Technology (INTECH), Fourth International Conference on IEEE, Luton., England*: 2014, 130-134.

[16] J. Chen, C.U. Zhao, C. Chen. Collaborative filtering recommendation algorithm based on user correlation and evolutionary clustering" *Complex & Intelligent Systems published online by Springer* 30 January, 2019, 147-156.

[17] H. Xia, Y. Luo, and Y. Liu. Attention Neural Collaboration Filtering Base on GRU for Recommender Systems." *Complex and Intelligent Systems. Published @ Springer.* 2020

[18] X. Liao, X. Li, Q. Xu, H. Wu, and Y. Wang. Improving Ant Collaborative Filtering on Sparsity via Dimension Reduction*" Applied Science*. 2021, 72-45 *http://www.mdpi.com/journal/applsci*

[19] J. McAuley, and J. Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. *In Proceedings of the 7th ACM conference on recommender systems.* 7(13): 2013, 165-172.

[20] P. Priyanga, and A. R. N. B. Kamal. Methods of Mining the Data from Big Data and Social Networks Based on Recommender System. *International Journal of Advanced Networking & Applications.* 8(5): 2017, 55-60.