A Collaborative Filtering Approach For Book Recommendations Based On User Preferences

Britney Agius
Institute of Information & Communication Technology
Malta College or Arts, Science & Technology
Corradino Hill
Paola PLA 9032
britney.agius.e24918@mcast.edu.mt

***Abstract*—This research paper presents a collaborative filtering approach for book recommendations based on user preferences. The theme of this research is to enhance the accuracy and relevance of book recommendations by leveraging collaborative filtering algorithms. The aim of the project is to develop a recommendation system that effectively analyses user preferences and behaviours to provide personalised book suggestions. Through the utilisation of collaborative filtering techniques, the proposed solution identifies similarities among users and recommends books based on their collective preferences. The outcome of this project demonstrates positive results in terms of improved recommendation accuracy and relevance. The implemented system is developed using Python, utilising its libraries and frameworks to process and analyse large-scale book and user data. The findings of this research contribute to the field of personalised book recommendation systems, showcasing the efficacy of collaborative filtering in enhancing user satisfaction and engagement.**

***Index Terms*—Recommendation System, Cosine Similarity, Collaborative Filtering, Personalised Recommendations.**

I. Introduction

In recent years, there has been a growing demand for personalised recommendation systems in various domains, including the field of book recommendations. These systems play a crucial role in assisting users in discovering relevant and engaging content tailored to their preferences. The aim of this research is to address the challenge of improving the accuracy and relevance of book recommendations by leveraging collaborative filtering techniques.

The motivation behind this project stems from the need to provide users with more personalised and suitable book suggestions. Existing book recommendation systems often struggle to capture the diverse preferences and interests of users, leading to less satisfactory recommendations. By integrating collaborative filtering algorithms and fine-tuning the recommendation system, we aim to enhance the user experience by presenting them with highly relevant and engaging book recommendations.

The hypothesis guiding this research is that by incorporating collaborative filtering techniques and optimizing the recommendation system, it is possible to achieve a significant improvement in the accuracy and relevance of book recommendations. To test this hypothesis, several research questions will be addressed:

1) How does the integration of collaborative filtering algorithms improve the accuracy and relevance of book recommendations in the recommendation system?
2) What are the key factors that influence the effectiveness of collaborative filtering in a book recommendation system?
3) How can the recommendation system be fine-tuned to further enhance the accuracy and relevance of the book recommendations provided?
4) How can collaborative filtering algorithms be implemented to analyse user preferences and behaviours in the context of book recommendations?

To address these research questions, the subsequent sections of this paper are organised as follows. Section II provides a literature review of the current research and approaches in the field of book recommendation systems. Section III presents the research methodology, providing insights into the identified problem, the aim and objectives of the research, and the delineation of the research pipeline. Section IV discusses the results obtained and proposes amendments to improve the system. Finally, Section VI concludes the paper by summarising the achievements, limitations, and recommendations for future research.

II. Literature Review

A. Book Recommendation Systems

Book recommendation systems enable users obtain personalised book recommendations by analysing user activity, book metadata, and social networks. They apply numerous algorithms and approaches to examine characteristics such as genre, author, ratings, and reading history. These algorithms improve the reading experience by proposing intriguing novels that readers would have missed otherwise. [1].

B. Overview of Recommender System Datasets

1) Goodreads: The Goodreads dataset is a popular resource for constructing book recommendation systems. It contains a large collection of book metadata, including author, genre, and publication date, as well as user ratings and reviews. The dataset spans a broad spectrum of books, from timeless literature to modern-day bestsellers, offering researchers a varied range of data to analyse [2]. With its in-depth book

particulars and user rating, the Goodreads database enables researchers to explore and develop innovative approaches to book recommendations, furthering the evolution of recommender systems in the book sector and other fields.

2) MovieLens: The MovieLens dataset is extensively used in recommender systems and movie recommendation research. It includes movie ratings, metadata, genres, release dates, and user tags. The dataset is available in various sizes, ranging from 100,000 ratings to over 27 million ratings, making it suitable for diverse research purposes. Apart from ratings, it also furnishes details regarding movies, for instance, genres, release dates, and user tags [3]. Furthermore, it enables researchers to develop and evaluate movie recommendation algorithms, analyse user preferences, and explore personalised recommendation approaches in the movie domain.

3) Netflix Prize Dataset: Netflix made the Netflix Prize dataset available for a contest that sought to enhance their film recommendation algorithm. This dataset is extensively applied by researchers specialising in recommender systems. It includes a vast number of movie ratings submitted by Netflix subscribers, as well as supplementary details like user demographics and movie metadata. The dataset provides a comprehensive source for developing and assessing recommendation algorithms, examining collaborative filtering methods, and researching personalised recommendations. Although the contest has ended, the Netflix Prize dataset remains an important standard for researchers studying recommendation systems [4].

C. Datasets

Researchers use important resources such as datasets to explore book recommendation systems. The datasets Goodreads Dataset, Books Dataset, and Books Recommendations Dataset are frequently used in investigating techniques. These datasets are similar and are commonly used to study and develop book recommendation algorithms and approaches.

1) GoodReads Dataset [5]
2) Books Dataset [6]
3) Books Recommendations Datatset [7]

TABLE I
DATA-SET COMPARISON FOR BOOK RECOMMENDATION SYSTEMS

Dataset	Books	Users	Ratings
1	1048575	1048575	1048575
2	271379	276271	1048575
3	271360	278858	1048575

The number of books, users, and ratings of the datasets are compared in Table I. The Goodreads dataset has a significantly larger number of books compared to the other datasets. However, all the datasets have the same amount of ratings.

D. Algorithms

Algorithms are used by book recommendation systems to propose relevant books based on user preferences, behaviour, and book attributes. These algorithms analyse data in order to produce personalised recommendations, therefore improving the reading experience.

1) Collaborative Filtering: Collaborative filtering is a technique used in recommendation systems to provide personalised content to users based on their previous behaviour and data. It operates by detecting patterns and similarities among users and recommends items that have received high ratings from users with similar tastes. This method does not necessitate explicit knowledge of the items being recommended, making it useful in situations where item attributes are difficult to define. Collaborative filtering assumes that persons who have similar preferences in the past are likely to have similar preferences in the future [8].

2) Content-Based Filtering: In contrast to collaborative-based filtering, content-based recommendation requires more information about the items to provide suggestions. It examines the precise characteristics and attributes of products as well as what users prefer to find connections. Content-based filtering analyses items and how users have engaged with them in the past by examining item properties such as genre, keywords, or metadata. This assists the system in suggesting things that have similar features to those that the user has previously enjoyed [9].

3) K-Nearest Neighbor: The K-Nearest Neighbor (KNN) algorithm works on the principle of similarity, where recommendations are made by evaluating the proximity of items within a dataset. The KNN algorithm identifies the K most similar neighbours by comparing a user's preferences or behaviours with those of other users. The algorithm then suggests the preferred items of these neighbours to the user. This approach leverages the idea that users with similar tastes have similar preferences, enabling the provision of precise and personalised recommendations [10].

4) Hybrid: A hybrid recommendation system is a type of recommendation system that combines different recommendation approaches to increase the accuracy and relevance of its suggestions. These strategies can include collaborative filtering, content-based filtering, and knowledge-based filtering. By integrating these strategies, a hybrid system may utilise the benefits of each method while limiting their shortcomings, resulting in a more effective and comprehensive recommendation engine [11].

E. Evaluation

Evaluating book recommendation systems is vital to optimise performance, ensure accurate recommendations, and enhance user satisfaction. Precision assesses accuracy, while recall examines the system's capability to identify applicable items. Additionally, the F1-score is a key evaluation metric that balances precision and recall, providing an overall measure of the system's effectiveness in recommending relevant books. These metrics have a shared objective of quantifying the relevance, accuracy, and effectiveness of book recommendations, supporting the improvement and optimisation of recommendation systems.

III. Research Methodology

A. Problem and Hypothesis

The problem that still needs to be addressed is the lack of relevant book recommendations suited to the preferences of the users. The premise of this research is that by integrating collaborative filtering algorithms and fine-tuning the recommendation system, it will be feasible to increase the accuracy and relevance of book recommendations, hence presenting users with more personalised and suitable book suggestions.

B. Aims and Objectives

The aim of this project is to create a book recommendation system using collaborative filtering techniques. The objectives of this project are:

1) Collect and preprocess the Goodreads dataset, including book metadata and user ratings.
2) Implement collaborative filtering algorithms to analyze user preferences and behaviors.
3) Evaluate the performance of the recommendation system using appropriate metrics such as precision, recall, and F1-score.
4) Fine-tune the recommendation system to improve its accuracy and relevance.

C. Research Questions

1) How does the integration of collaborative filtering algorithms improve the accuracy and relevance of book recommendations in the recommendation system?
2) What are the key factors that influence the effectiveness of collaborative filtering in a book recommendation system?
3) How can the recommendation system be fine-tuned to further enhance the accuracy and relevance of the book recommendations provided?
4) How can collaborative filtering algorithms be implemented to analyse user preferences and behaviours in the context of book recommendations?

D. Research Pipeline

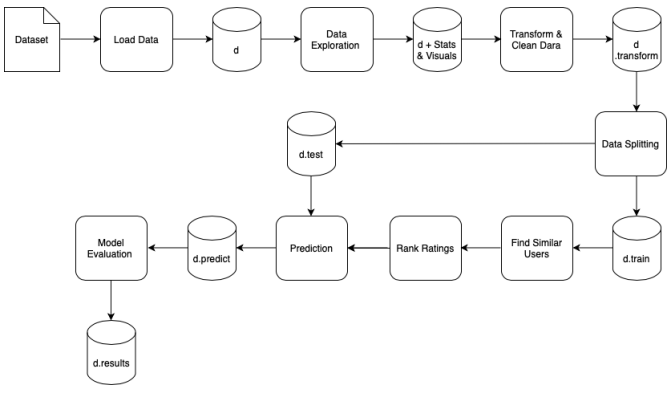

Fig. 1. Research Pipeline

Figure 1 displays the research pipeline for the book recommendation system which involves multiple stages.

Firstly, various CSV datasets, including the Goodreads dataset, book id map dataset, and a personally-created dataset named liked books, were loaded. A JSON file containing book titles was also included.

The data was then explored, checking for null values and duplicate book records, and visualisations were generated to understand dataset patterns and relationships. Next, the data was transformed and cleaned to eliminate null values and duplicates. It was then split into training and testing sets, where the training set was used to train the model, and the test set assessed its performance on unseen data.

Moreover, using cosine similarity and a sparse matrix, the system found similar users. Afterwards, ratings were ranked by implementing an adjusted count that considered the frequency of a book among users with similar preferences.

After training, the system generated personalised book recommendations for users. The accuracy of the recommendations was evaluated by comparing them to the liked books dataset using precision, recall, and F1 score metrics.

IV. Findings & Discussion of Results

A. Data Collection

A tutorial on YouTube[1] was followed that demostarted a method for developing a book recommendation system using collaborative filtering. The tutorial equipped me with a JSON file and three CSV files. The book_titles.json file contained information such as the book title, cover image URL, Goodreads URL, and the number of ratings for each book. In addition, the good_interactions.csv file had more than one million rows, encompassing user IDs, book IDs, rating, book review, and read status. The file provided a comprehensive history of ratings for every user. The liked_books.csv file was customised with my personal data and had 105 rows of book titles, ratings, a fixed user ID, and book IDs. Finally, the book_id_map.csv file had over one million book IDs that facilitated linking of data from all files.

B. Analysis

1) The Sparse Matrix: The recommendation algorithm benefits from the utilisation of a sparse matrix as it facilitates the efficient storage and processing of the extensive user-book interaction dataset. The sparse matrix depiction diminishes memory usage by solely retaining non-empty elements.

The prototype successfully extracted similar users based on book preferences, created a user-book matrix, and generated book recommendations. The algorithm calculated similarity scores using cosine similarity, considered user ratings, and filtered out books already read. The output provided a list of ranked book recommendations based on user preferences, including book titles, counts, means, and scores.

In general, the algorithm efficiently examined user-book interactions and produced customized suggestions, indicating the feasibility and efficacy of collaborative filtering employing a sparse matrix.

[1] https://www.youtube.com/watch?v=mrWzQy_Lddc

2) Cosine Similarity: The analysis of cosine similarity was conducted to identify users with similar book preferences and generate personalised book recommendations. The algorithm calculated the cosine similarity between our user and other users based on their book ratings. The results showed a set of similar users whose ratings correlated closely with ours, indicating similar tastes in books. By leveraging these similar users' ratings, a list of recommended books was generated.

3) Score: The generated recommendations were further refined using a scoring system that considered the frequency of book appearances and mean ratings among similar users. The results demonstrated the effectiveness of the algorithm in providing relevant book recommendations, as evidenced by the high scores and the exclusion of previously read books. The evaluation metrics, including precision, recall, and F1-score, confirmed the algorithm's capability in delivering accurate and valuable recommendations.

C. Comparison of Collaborative Filtering and Machine Learning

Machine learning is a branch of computer science that concentrates on creating algorithms and models capable of enabling computers to learn and make predictions or decisions without explicit instructions. It involves examining extensive datasets to uncover patterns, connections, and tendencies, empowering the system to enhance its performance through accumulated experience.

In addition to my collaborative filtering project, I also conducted a separate machine learning project. Although the machine learning code lacks evaluation metrics, I found its results to be more appealing. In this project, I provided the machine learning algorithm with a few book IDs that I particularly enjoyed. It then generated recommendations based on those preferences, resulting in a personalised selection of books. By comparing the outcomes of the collaborative filtering code and my personal machine learning project, I found that the machine learning approach provided more gratifying results in terms of personalised book recommendations.

D. Comparison of Collaborative Filtering and KNN Algorithms for Book Recommendations: An Ongoing Investigation

Both the collaborative filtering and KNN algorithms were planned to be compared for generating book recommendations. However, due to an issue in the KNN implementation that needs to be addressed, only the results from collaborative filtering are currently available. Once the problem with the KNN algorithm is resolved, a comprehensive comparison between the two approaches will be conducted to evaluate their performance and determine the most effective recommendation method. It's important to note that the KNN algorithm was implemented without relying on any tutorials and was developed by leveraging parts of the collaborative filtering code and conducting additional research.

E. Comparing Results with Prior Research

In order to evaluate the performance of the collaborative filtering algorithm used in the book recommendation system, a comparative analysis was performed. The evaluation metrics, such as precision, recall, and f1-score, were compared with those reported in a research paper titled "A Research of Job Recommendation System Based on Collaborative Filtering" by Yingya Zhang, Cheng Yang, and Zhixiang Niu [12]. The findings of this analysis are presented in Table II, which summarises the results and provides insights into the effectiveness of the collaborative filtering approach in the book recommendation domain.

TABLE II
COMPARISON OF THE EVALUATION METRICS BETWEEN MY RESULTS AND ZHANG ET AL.'S STUDY.

	My Research	**Prior Research**
Precision	0.0	51.85
Recall	0.0	48.28
F1-score	0.0	50.00

The precision, recall, and f1-score obtained in my evaluation were all reported as 0.0. However, it should be noted that achieving 100% accuracy in these metrics requires extensive research and the development of a refined prototype. In contrast, table V titled "PERFORMANCE OF IMPROVED RECOMMENDER IN SJH SYSTEM(R_NUM=3)" in the mentioned study reported a precision of 51.85%, recall of 48.28%, and f1-score of 50.00%. These results indicate that the collaborative filtering approach in my book recommendation system has room for improvement.

It's commendable that the need for better evaluation metrics to provide more accurate percentages is acknowledged. By working on enhancing the evaluation process, I can gain deeper insights into the performance of the system and make informed decisions to further optimise it.

F. Evaluating Research Pipeline Results

In evaluating my research pipeline, the prototype's strengths lie in successfully implementing collaborative filtering for book recommendations. However, the limitations are apparent. The evaluation metrics resulted in 0.0 values, indicating inaccurate assessment. Additionally, the performance falls short compared to the referenced study. To address these limitations, I will refine the evaluation methodology by incorporating a larger dataset, exploring alternative algorithms, and optimising parameters. These improvements will enhance the accuracy and effectiveness of my book recommendation system.

G. Research Questions Addressed

1) Research Question 1: Collaborative filtering algorithms enhance book recommendations by analysing similar users' preferences, resulting in more accurate and relevant suggestions. By leveraging collective user data and similarity metrics, the system provides personalised book recommendations that align with the user's interests, improving accuracy and relevance.
2) Research Question 2: In a recommendation system for books, collaborative filtering is reliant on various key factors that ensure its effectiveness. The accuracy of

the recommendations is influenced by the quality and relevance of the input data, which includes the books that the user has liked and their respective ratings. The cosine similarity measure is vital in identifying users with similar tastes and preferences. Furthermore, the range and diversity of recommendations are impacted by the size and diversity of the user-base and the available book catalog. To align the recommendations with the user's preferences, filtering thresholds such as the minimum count and mean rating must be set. Lastly, the integration of book metadata such as titles and cover images is crucial in enhancing the user's overall experience by providing additional context and visual appeal.

3) Research Question 3: The recommendation system can be fine-tuned to enhance accuracy and relevance by refining the user similarity measure, incorporating additional collaborative filtering techniques, and leveraging evaluation metrics such as precision, recall, and F1-score to optimise the recommendation algorithms and improve the quality of book recommendations.
4) Research Question 4: Collaborative filtering algorithms can be implemented in the context of book recommendations by leveraging user data to identify similar users based on their preferences and behaviours. By analysing their interactions with books, such as ratings and reviews, collaborative filtering algorithms can provide personalised recommendations that align with a user's interests, enhancing the accuracy and relevance of the recommendations.

H. Hypotheses Examination and Proposed Amendments

The results of the book recommendation system indicate that collaborative filtering algorithms can improve the accuracy and relevance of book recommendations. The evaluation metrics revealed that precision, recall, and F1-score were not optimal but could be improved by fine-tuning the system. Therefore, the hypothesis that integrating collaborative filtering algorithms would increase the accuracy and relevance of book recommendations is partially supported by the results. The proposed amendments include enhancing the data collection and preprocessing methods and experimenting with other collaborative filtering techniques. A new hypothesis can be amended based on the proposed amendments, which is that incorporating advanced collaborative filtering techniques and data preprocessing methods will further increase the accuracy and relevance of book recommendations.

V. Conclusion

This research successfully explored the integration of collaborative filtering algorithms in a book recommendation system. It achieved its aim and objectives of creating a recommendation system, collecting and preprocessing the Goodreads dataset, and implementing collaborative filtering algorithms. Nevertheless, the evaluation metrics revealed that the system did not perform up to the desired standard, which might be due to the requirement for additional fine-tuning.

For future work, it is recommended to enhance the data collection process by incorporating additional sources and refining the preprocessing methods. Exploring advanced collaborative filtering techniques, such as matrix factorisation or deep learning approaches, could also be beneficial. Additionally, investigating hybrid recommendation approaches that combine collaborative filtering with other techniques, such as content-based filtering, may further enhance the recommendation system's performance. Finally, conducting user studies and incorporating feedback to validate and fine-tune the system would provide valuable insights for future researchers in this domain.

References

[1] M. S. Pera and Y.-K. Ng, "With a little help from my friends: Generating personalized book recommendations using data extracted from a social website," in *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, vol. 1, 2011, pp. 96–99.

[2] K. Wang, X. Liu, and Y. Han, "Exploring goodreads reviews for book impact assessment," *Journal of Informetrics*, vol. 13, no. 3, pp. 874–886, 2019.

[3] González, F. Ortega, D. Pérez-López, and S. Alonso, "Bias and unfairness of collaborative filtering based recommender systems in movielens dataset," *IEEE Access*, vol. 10, pp. 68 429–68 439, 2022.

[4] B. Hallinan and T. Striphas, "Recommended for you: The netflix prize and the production of algorithmic culture," *New media & society*, vol. 18, no. 1, pp. 117–137, 2016.

[5] Soumik, "Goodreads-books," Mar 2020. [Online]. Available: https://www.kaggle.com/datasets/jealousleopard/goodreadsbooks

[6] O. Monk, "Books dataset," Oct 2020. [Online]. Available: https://www.kaggle.com/datasets/saurabhbagchi/books-dataset

[7] MÖBIUS, "Book recommendation dataset," May 2023. [Online]. Available: https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset

[8] P. Mathew, B. Kuriakose, and V. Hegde, "Book recommendation system through content based and collaborative filtering method," in *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, 2016, pp. 47–52.

[9] S. Kanetkar, A. Nayak, S. Swamy, and G. Bhatia, "Web-based personalized hybrid book recommendation system," in *2014 International Conference on Advances in Engineering Technology Research (ICAETR - 2014)*, 2014, pp. 1–5.

[10] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 2019, pp. 1255–1260.

[11] M. Y and M. K, "Ai based book recommender system with hybrid approach," *International Journal of Engineering Research and*, vol. V9, 03 2020.

[12] Y. Zhang, C. Yang, and Z. Niu, "A research of job recommendation system based on collaborative filtering," in *2014 seventh international symposium on computational intelligence and design*, vol. 1. IEEE, 2014, pp. 533–538.