A Hybrid Book Recommendation System for University Library

Pitiwat Arunruviwat
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Bangkok, Thailand
6272053521@student.chula.ac.th

Veera Muangsin
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Bangkok, Thailand
veera.m@chula.ac.th

2022 26th International Computer Science and Engineering Conference (ICSEC) | 978-1-6654-9198-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSEC56337.2022.10049318

***Abstract*—Nowadays, Recommendation systems become an important role in daily life such as recommended goods, recommended musics, recommended books, or recommended movies. Furthermore, a university library initiated a book recommendation system for improving the efficiency of book searching. This paper presents a methodology for book recommendation in a university library using a hybrid recommendation technique by weighting a combination of 2 similarity scores from two recommendations system. Normally a hybrid recommendation system is built on a combination of content-based filtering and collaborative filtering, whereas this paper use technique for applied books from the Course Syllabus and combines it with a standard hybrid recommendation system. To solve the cold start problem and improve the accuracy of the book recommendation system in the university library. For the evaluation, RSME has been used to collaborate with K-Fold Cross Validation and Train Test Split technique. Eventually, the result of the evaluated book recommendation system shown RSME is 1.2061 for 5-Fold Cross Validation and 1.2247 for Train & Test Split**

Keywords—Recommendation, Content-based Filtering, Collaborative Filtering, Hybrid Filtering

I. INTRODUCTION

In technological world, personal recommendation systems have been introduced. It is used to create recommendations for each user, such as recommendations for songs products, books or movies, etc. For an example website where the recommendation system has been applied is Amazon book selling (Kindle). The kindle has several book recommendations, such as book recommendation based on book that users have bought, new books, popular books, etc. and when the user buy a book, The system will recommend the next book that the user is expected to be interested as well. Video streaming websites such as Netflix provide movie recommendations to each user based on genre that user preferred and movies similarity to user has seen before.

There are 3 techniques used to recommend books based on user's interests. 1) content-based Filtering 2) Collaborative Filtering 3) Hybrid Filtering. One of most popular technique for books recommendation is collaborative filtering. However, this technique is limitation with cold start problem

This paper has an idea to apply the recommendation system to university libraries combine with book from Course Syllabus. Because the [1] university library is a library that serves to support studying, learning according to the curriculum and researching both teachers and students. But books in libraries have been continuously increasing so fast. As a result, users face the problems in finding books that do not meet user's satisfaction and takes a long time to searching books in the library. In addition, the library system still lacks in the part of the book recommendation system that will help users when finding books. Combining information from [2] Course Syllabus that contain an academic document that communicates about a specific course such as course name, course code and course description which is a path of data for this paper to be used in hybrid recommendation system.

As previously mentioned, this paper aims to use information from books in university library and Course Syllabus to applied and build hybrid recommendation system based on content-based filtering combine with collaborative filtering to achieve the efficiency and effective results of books recommendation.

The rest of this paper is organized as follows. Section II discuss about related work on using varies recommendation system techniques in area of book recommendation system, hybrid book recommendation system and university library recommendation system. In Section III, we describe the method for preparing books data, build recommendation system and combining recommendation system. In Section IV, we present the results and an evaluation of models, and finally, we summarize our paper in section V

II. RELATED WORK

A. Book Recommendation System

Darekar et al. [3] use bestselling books purchased history and books title to calculate and predict using TF-IDF technique. The results show that the title of the book that matches bestselling purchased history is likely to become a popular book

Nursultan Kurmashov & Konstantin Latuta [4] build a book recommendation system using a collaboration filtering technique. Datasource from user give a rating score and favorite book categories

In an aspect of book recommendation system, using a single technique of recommendation system might not be effective enough and the result may not meet user need

B. Hybrid Recommendation System

Darekar et al. [5] use 3 techniques of recommendation system (collaborative filtering, content-based filtering, and demographic filtering) to build a hybrid recommendation system. They compare results only if exist 3 techniques of recommendation system. The result shows that using a hybrid recommendation system gave accuracy and efficiency to user expectations more than using only 1 technique of the recommendation system

Yang et al. [6] gathering data from the user's access history, user rating, and item detail description to predict with collaborative filtering. Then evaluate with MAE (Mean Absolute Error). The result has shown that using hybrid recommendations based on user information gave more efficiency and effectiveness

Hong-Quan et al. [7] focus on weighted hybridization rather than using fixed weight for the combination of 2 recommendation systems (content-based and collaborative). The experimental results on one of the most popular public datasets in the field of RSs - MovieLens have verified the effectiveness of our strategy against the traditional CF, and CBF approaches. It not only boosts the prediction performance but also alleviates the problem of new item cold star

In an aspect of hybrid book recommendation system, the result from related work proved that using hybrid recommendation systems help Increased efficiency of recommendation system and meet users' expectation greater than using of any one recommendation system

C. *University Library Recommendation System*

Yonghong et al. [8] use collaborative filtering techniques combined with content-based filtering to solve the data sparsity since the information in the library is large, but the number of borrowing users may not be diverse. K-means clustering had been applied to help solve data sparsity. The result shows that applied 2 techniques help increase accuracy and precision compared with using only 1 technique

Guang Liu & Xu Zhao [9] propose a model that transfers the implicit data of readers borrow history to explicit data and applies the SVD++ algorithm in the recommender system. The model takes the natural logarithm of the total borrowing days of the reader as the reader's rating, to make an automatic grading for the reader's preference. Since ln1=0, the model takes the length of borrowing time is 1 day as negative feedback and takes the length of borrowing time over 1 day as positive feedback

Li et al. [10] proposes an interest-based university book recommendation algorithm. The algorithm solves the problem of the lack of scoring and the inability to use the collaborative filtering algorithm. Datasets from borrowing records, book searching records, and the number of borrowed books for calculating similarity score instead of using user raring. Experiments show that the algorithm is better than the traditional collaborative filtering recommendation algorithm and has certain recommendation effects and practical value in the application of university libraries

In an aspect of the university library recommendation system, each recommendation techniques have advantages in different applications in different contexts of use. at present, the collaborative filtering recommendation system is still the most popular. But the important is datasets must be enough to use to predict the recommendation system. Otherwise, the problems with fewer datasets will lead to low accuracy prediction and not enough data variety (Sparseness)

III. METHODOLOGY

The methodology was divided into 6 sections. 1) gathering data from university library, 2) gathering data from Course Syllabus, 3) data preparation to build hybrid recommendation system, 4) construct content-based recommendation 5) construct collaborative filtering, and 6) construct hybrid recommendation system using weighting score technique

A. *Gathering Data from University Library*

Datasets in the university library came from the database Sierra Library Information systems which examples are shown in Table I and Table II. The datasets have total of 139,214 records exported from the database

TABLE I. EXAMPLE SIERRA INFORMATION SYSTEM

Column	Value
Bib_record_id	420909003711
Title	Buddhism in China
Author	Ling Haicheng
Publish_year	2004

TABLE II. EXAMPLE BOOK BORROWING RECORD

Name	Description
Bib_record_id	450973900101
Item_record_metadata_id	420909003711
Patron_record_metadata_id	481037631869
Checkout_gmt	03/01/2018
Rating*	1-5

Since there is no explicit rating for each book, it is necessary to use an implicit rating based on the borrowing book count, categories of members in the library, and period of borrowing book of each member. To give a rating, the idea is if borrowers in the same category borrow more books (Calculated by a proportion of the number of latest borrowing counts in the member's category divided by the total number of members in the group who borrow books). It means that the book is more likely to be interesting to members of that category. Therefore, giving a high rating for borrowing by members of the same category in Table III

TABLE III. EXAMPLE OF BORROWING BOOK IN FACULTY OF COMMUNICATION ARTS

Faculty/Book	Borrowing Count	Latest Borrowing Date	Percentage	Rating
The Odyssey	1	11/08/18	0.01%	1
Cambridge grammar of English	1	11/08/18	0.01%	1
Microsoft Word 2017	1	11/08/18	0.01%	1
Theory and Algebra	18	06/08/18	0.18%	5
แคลคูลัส: เล่ม 4	19	25/05/21	0.19%	5
ทฤษฎีจำนวน	20	06/08/21	0.20%	5
เก็นเจ๊าหนูสู่ชีวิต	39	06/08/21	0.39%	5

จิตวิทยาการศึกษา	45	25/05/21	0.45%	5

Table III shows an example of the number of books borrowed for each book in Faculty of Communication Arts students because there is the greatest number of borrowing books in this faculty and most students in this faculty borrowed books from the university library, in order from least to the greatest number of books borrowed and from longest to most recent borrowing period. There are 10,362 students in the Faculty of Communication Arts who have all book borrowing data in the Sierra Library Information System. Then used the number of books borrowed in each book was divided by total number of students in the Faculty of Communication Arts who borrowed books. To calculate a percentage of the ratio and summarize the results of the percentage of the ratio for calculating the rating score by specifying the method of dividing the rating score as in Table IV

TABLE IV. RATIING CALCULATION BASED ON SUMMARIZE OF RATION

PERCENTAGE RATIO	RATING
0-20%	1
21-40%	2
41-60%	3
60-80%	4
80-100%	5

B. *Gathering Data from Course Syllabus*

This paper use datasets from Course Syllabus of faculty of engineering which example show as Table V. The datasets have total 553 records exported in CSV format

TABLE V. EXAMPLE COURSE SYLLABUS IN FACULTY OF ENGINEERING

Name	Description
Course No	2301457
Course Name	MATH PERSP ML ALGO
Material Type	Recommended Book
Title	Mathematics for Machine Learning
Author	Marc et al.
ISBN	978-1108455145
Description*	Book Description

Since there is no book description in Course Syllabus, it is necessary to implement API to scrape book descriptions from Google API

C. *Data Preparation*

Data preparation is a very important process because it is a process that directly affects the accuracy score of the recommendation model and also affects to similarity calculation

- Removal of special characters using regular expression, such as (,.!$&()-=@{}[]-]") etc. will be removed in this step
- Removal of stop words in every sentence, for the similarity score calculation, it is necessary to remove words that are not highly significant
- Word Tokenization, it is importance to tokenize each word in the sentences using tokenization library

D. *Content-based Filtering*

Building content-based filtering is the first step of hybrid recommendation. This step use student information (ID, Name, Faculty) along with datasets from Course Syllabus especially book name and description to calculate the similarity score by using TF-IDF technique and cosine similarity. The process described in Fig. 1

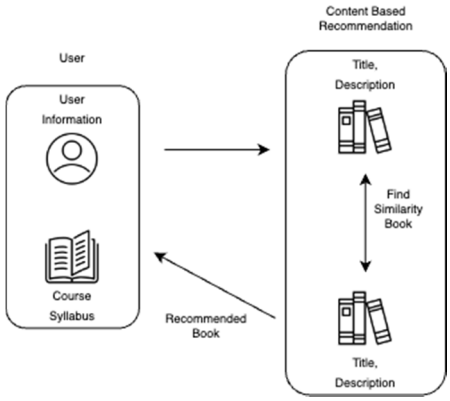

Fig. 1. Content base Filtering

E. *Collaborative Filtering*

Building Collaborative filtering is the second step of hybrid recommendation. This step use student information (ID, Name, borrowing/returning history) along with the user's score on each book after the user has borrowed/returned the book. Then, the recommendation system will use the scores from each book to calculate the Pearson Correlation Coefficient [11] with neighboring books and sorted the score from highest to lowest. The process described in Fig. 2

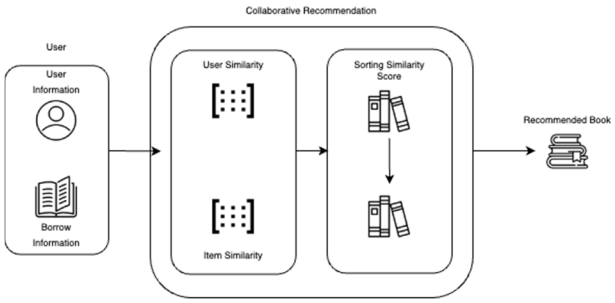

Fig. 2. Collaborative Filtering

F. *Contstruct Hybrid Recommendation*

This step is to apply the results of the content-based book recommendation system and the collaborative book recommendation system. Combining the similar score from both recommendation systems and calculate the average weight of scores in Fig. 4. Then the recommendation system sorted the score by using the weighing technique. Scores from highest to lowest will display as a list of book recommendations. The process described in Fig. 3

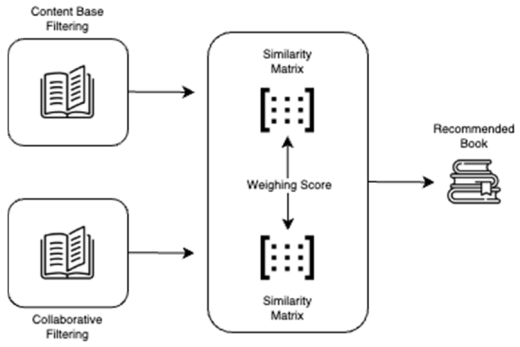

Fig. 3. Hybrid Recommendation

item	rec_1 score	rec_1 rank	rec_2 score	rec_2 rank	rec_w score	rec_w rank
Item1	0.5	1	0.8	2	0.65	1
Item2	0		0.9	1	0.45	2
Item3	0.3	2	0.4	3	0.35	3
Item4	0.1	3	0		0.05	
Item5	0		0		0	

Fig. 4. Example of Weighing Calculation in Hybrid Recommendation

G. Evaluation Method

For evaluating the effectiveness of the recommendation system, it is divided into two parts. (1) Using the Root Mean Square Error (RMSE) equation to help evaluate the results of the recommendation system (2) Using examples users to try out book searching system

IV. Result and Evaluation

A. Hybrid Recommendation System

As described in the section earlier, the result of hybrid recommendation system shown in table VII and the result of content-based filtering and collaborative filtering shown in table VI and table VII with example of searching book "Applied Mining Geology"

TABLE VI. EXAMPLE RESULT OF CONTENT-BASED RECOMMENDATION

Title	Score
Introduction to Mineral Exploration	0.2276
Numerical Methods in Engineering	0.1268
Pattern Recognition and Machine Learning	0.1051
Environmental Toxicology	0.1017

TABLE VII. EXAMPLE RESULT OF COLLABORATIVE RECOMMENDATION

Title	Score
Introduction to Data Mining	1.0

TABLE VIII. EXAMPLE RESULT OF HYBRID RECOMMENDATION

Title	Score
Introduction to Data Mining	0.500
Introduction to Mineral Exploration	0.0852
Numerical Methods in Engineering	0.0634
Pattern Recognition and Machine Learning	0.0525
Environmental Toxicology	0.0508

B. Evaluation

The paper used users' data, books and ratings to measure the efficiency of the hybrid book recommendation system by using the Root Mean Square Error (RMSE) equation with the K Fold Cross Validation from Surprise, a Python scikit for evaluation of book recommendation system performance, it divided the data into five parts (5-fold) before applying to experiment with different algorithms as follows:

- Normal Predictor means an algorithm predicting a random rating based on the distribution of the training set, which is assumed to be normal
- Baseline Only means an algorithm predicting the baseline estimate for given user and item
- SVD means an algorithm that classifies items and users according to multiple factors. Inferences from the rating scoring model
- KNNBasic means an algorithm for calculating the relationship between items or between neighboring users based on the ratings of the user's items

To calculate the lowest possible RSME obtained by performing 5 Fold Cross Validation from all above algorithms. It can be shown in Table IX

TABLE IX. RESULT OF 5 FOLD CROSS VALIDATION FROM ALL ALGORITHM

Algorithm	RMSE	Fit Time	Test Time
Baseline			
1 Fold	1.2802	0.3140	0.0670
2 Fold	1.2822	0.4610	0.2108
3 Fold	1.2823	0.3423	0.0666
4 Fold	1.2745	0.4731	0.0646
5 Fold	1.2885	0.5088	0.0686
Normal Predictor			
1 Fold	1.9329	0.0698	0.0854
2 Fold	1.9184	0.0781	0.0849
3 Fold	1.9409	0.0875	0.0733
4 Fold	1.9295	0.0778	0.0722
5 Fold	1.9280	0.0800	0.0783
SVD			
1 Fold	1.2686	3.3569	0.1470
2 Fold	1.2567	2.2713	0.1437
3 Fold	1.2787	1.8465	0.3508
4 Fold	1.2651	1.6881	0.1312
5 Fold	1.2695	1.6428	0.1276
KNN			
1 Fold	1.2049	3.4352	0.0950
2 Fold	1.2105	3.1941	0.0814
3 Fold	1.2062	3.2263	0.0959
4 Fold	1.2043	3.2720	0.0750
5 Fold	1.2027	3.3856	0.1926

From Table IX, the total average 5 fold of RSME calculated from KNN is 1.2057 which is the smallest of the total mean of all algorithms. Therefore, the KNN algorithm will be used to meature the predictive accurary of the recommendation system. Futhermore, This paper used train & test split technique from surprise lib in Python to sampling training data and testing data in ratio 70:30 and using RSME

to predictive acccuracy of the recommendation. The result of RSME Train & Test split shown as Table X

TABLE X. RESULT OF TRAIN & TEST SPLIT TECHNIQUE

Algorithm	Train Data	Test Data	RMSE
KNNBasic	70%	30%	1.2247

In order to prove the hybrid book recommendation proposed by the author, Conducting an experimental comparison of our technique with other techniques (content-based filtering, collaborative filtering). The RMSE are calculated based on books data in Section III and used the same Train & Test split both algorithm and ratio as Table X in section IV. The experimental comparison result shown as Table XI

TABLE XI. AN EXPERIMENTAL TECHNIQUE COMPARISON

Technique	RSME
Contented-Base Filtering	1.6945
Collaborative Filtering	1.4137
Our Hybrid Filtering	1.2247

It can be concluded that our hybrid recommendation technique with the applied book from the Course Syllabus gave more accurate than other techniques because lower RSME means a more accurate the prediction model

V. CONCLUSION

In this study, we presented a method for creating a hybrid book recommendation system within a university library using information from the book library and Course Syllabus. Hybrid Filtering with weighing technique have been applied to this study. The datasets based on combination of content-based filtering and collaborative filtering to get the results of the book recommendation and improving accurary of the book reccomendation. Finally, RSME has been used to collaborate with K-Fold Cross Validation and Train Test Split technique. Eventually the result of the evaluated book recommendation system shown RSME is 1.2061 for 5-Fold Cross Validation and 1.2247 for Train & Test Split.

The effectiveness of the book recommendation system and the book similarity score may differ due to the effectiveness of the recommendation technique and the information used in the book recommendation system

REFERENCES

[1] Encyclopedia of Library and Information Sciences, Third Edition

[2] “Recommender Systems Handbook” Springer Science & Business Media, pp. 305

[3] Sarun Juntui, Paween Khoenkaw: Automatic non-personalized Book Recommender Algorithm for Bookstore Shelf Management, in the 3rd International Conference on Digital Arts, Media and Technology (ICDAMT 2018)

[4] Nursultan Kurmashov, Konstantin Latuta: Online Book Recommendation System, in 2015 Twelve International Conference on Electronics Computer and Computation (ICECCO 2015)

[5] Rohit Darekar, Karan Dayma, Rohan Parabh: A Hybrid Model for Book Recommendation, in International Conference on Inventive Communication and Computational Technologies (ICICCT 2018)

[6] Junrui Yang, Cai Yang, Xiaowei Hu: A Study of Hybrid Recommendation Algorithm Based on User, in 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics

[7] Hong-Quan Do, Tuan-Hiep Le, Byeongnam Yoon: Dynamic Weighted Hybrid Recommender Systems, in International Conference on Advanced Communications Technology (ICACT 2020)

[8] Yonghong Tian, Bing Zheng, Yanfang Wang, Yue Zhang, Qi Wu, College Library Personalized Recommendation System Based on Hybrid Recommendation Algorithm, in 11th CIRP Conference on Industrial Product- Service Systems 2019

[9] Guang Liu, Xu Zhao: Recommender System for Books in University Library with Implicit Data, in International Conference on Network, Communication, Computer Engineering (NCCE 2018)

[10] Jiabei Li, Tian wei Xu, 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 (CSE 2020)

[11] “Recommender Systems and the Social Web: Leveraging Tagging Data for Recommender Systems”, Springer Science & Business Media, pp. 10