
A Context-Aware Personalized Hybrid Book Recommender System

Hossein Arabi¹, Vimala Balakrishnan^{2,*} and Nor Liyana Mohd Shuib²

¹*Bookurve Sdn. Bhd., Malaysia*

²*Department of Information Systems, University of Malaya, Kuala Lumpur 50603, Malaysia*

E-mail: hossein.arabi@gmail.com; vimala.balakrishnan@um.edu.my;

liyanashuib@um.edu.my

**Corresponding Author*

Received 19 December 2018; Accepted 20 April 2020;
Publication 14 July 2020

Abstract

Contextual information such as emotion, location and time can effectively improve product or service recommendations, however, studies incorporating them are lacking. This paper presents a context-aware recommender system, personalized based on several user characteristics and product features. The recommender system which was customized to recommend books, was aptly named as a Context-Aware Personalized Hybrid Book Recommender System, which utilized users' personality traits, demographic details, location, review sentiments and purchase reasons to generate personalized recommendations. Users' personality traits were determined using the Ten Item Personality Inventory. The results show an improved recommendation accuracy compared to the existing algorithms, and thus indicating that the integration of several filtering techniques along with specific contextual information greatly improves recommendations.

Keywords: Recommendation system, context – aware, personality, demographic, location, review sentiment, purchase reason.

Journal of Web Engineering, Vol. 19_3-4, 405–428.

doi: 10.13052/jwe1540-9589.19343

© 2020 River Publishers

1 Introduction

Recommendation Systems (RSs) are programs, software techniques and methods suggesting items to be of interest to a user. RSs are moving towards people who need personal assistance in making decision choosing relevant items among overwhelming number of alternative items that a system or website may offer, hence the term personalized RSs. Previous studies on personalized RSs show that by utilizing users' personality characteristics, for example, a higher recommendation accuracy can be achieved [1].

Existing studies have also focused on several contextual features to provide improved personalized recommendations, such as emotion or mood to recommend movies, music [2, 3], and user locations and preferences for travel recommendations [2, 4], etc. As users of RSs may have different needs in various situations and contexts, it is becoming increasingly important to consider contextual data when filtering information.

The use of contextual features for personalized recommendations has showed promising results, however, studies integrating these features to improve recommendation accuracy are scarce. In fact, existing studies tend to favor a specific domain for identified features, for example, mood, time and emotion are often explored in music/movie domain, whereas location in tourism. Contextual features need not be domain specific as shown by Cantador et al. [5], where correlations were found between users' personality traits across multiple domains. For example, users who scored high on Openness (i.e. personality) prefer educational books and country music, suggesting that a feature that is used to determine users' preferences for product A can be used to determine their preferences for product B.

Aside from users' contextual features, product features such as reviews, description, etc. play important roles in improving recommendations as well [6]. However, the textual information obtained from a product's features is usually very short or involves language ambiguity that makes it difficult to estimate users' interest correctly. Recent studies have begun utilizing sentiment analysis and Natural Language Processing tools to perform analysis on the acquired textual information to improve recommendations.

The review shows that despite the number of studies on personalized RSs, a lot more can be accomplished to improve the recommendations [7, 8]. The current study aims to extend the literature by addressing the above-mentioned gaps by proposing a personalized RS that integrates several user and product contextual features in order to improve the recommendation accuracy. The RS was tailored to support book recommendations; hence, it

is aptly named as Context-Aware Personalized Hybrid Book Recommender (hereinafter referred to as CAPHyBR).

This study is an extension of our previous study [9], on developing a personalized book recommender by integrating users' contextual features (i.e. demographic, personality trait and location). In the current study, the proposed RS enhanced the previous work by considering the products' features (review semantic analysis and purchase reason) together with users' contextual features. The effectiveness of product features was measured and compared with other existing personalized RSs. The rest of this research article is structured as follows: Section 2 presents some of the related studies, followed by the research design and methodology in Section 3. Section 4 provides the experimental results and discussions, and the paper is finally concluded in Section 5.

2 Literature Review

RSs are tools and suggestion methods for products or items for the users. "Item" is the term used to determine what is being recommended to users by the system. An RS's focus is on a particular type of item (such as movies, books, music or news) and its planning, its graphical user interface, and the main suggestion method to prompt the recommendations, are all made to provide effective and meaningful recommendations for a particular product [7].

Earlier studies on users' personality support the possibility of using personality information in RSs. It is meaningful to differentiate each users based on their personality that can be used in a very wide range of RSs. For instance, a user's interest in music is very relative with his/her personality type [10].

The most comprehensive and popular method to determine personality is based on the Five Factor Model or the Big Five model [11], which is a hierarchical organization of personality traits in terms of five basic dimensions: extraversion (E), agreeableness (A), conscientiousness (C), neuroticism (N) and openness (O). Big Five is considered to represent the basic dimensions of user personality, as its dimensions are steady, cross-culturally applicable and have biological basis [11]. It is also one of the most widely used and recognized instruments in determining a user's personality [4]. There are many alternative tests to measure Big Five's dimensions, one of which is the Ten Item Personality Inventory (TIPI), a popular test containing only 10 questions to effectively and efficiently calculate the five dimensions of personality [12].

It is well known that human reasoning and decision-making are strongly influenced by psychological aspects. Recent works have explored the adoption of personality traits to provide personalized recommendations. For instance, a more recent study used users' Facebook profiles to determine their personality traits in order to help improve book recommendations [1].

Collaborative Filtering (CF) is one of the most widely used technique that utilizes a user's info (such as personality trait, demographic, behavior and etc.) to provide more personalized recommendations [13]. CF stores and analyzes the user's info and tries to find a set of similar user behavior to generate recommendations. In other words, it recommends items that people with similar tastes and preferences liked in the past. CF has been used widely in social networking studies [14] and also book recommendations [15].

The main aim of using CF approach is to generate customized recommendations based on user preferences and interests, therefore it can be advantageous to incorporate several contextual features in the recommendation process. For instance, Choi et al. [16] proved that by adding social factors to the RS, they could gain the user's trust leading to a better performance.

Similarly demographic information such as gender and age has been used to find similar groups of people, and thus improving recommendations [17]. An early RS, referred to as Grundy, recommended books based on the library of manually assembled stereotypes [18]. The characteristics of the users were collected with the use of an interactive dialogue, which showed an improved recommendation. However, studies using demographic filtering in personalised RSs are completely lacking.

In addition to items and users, the RS needs to consider their context or environment attributes while generating the list of recommendations. As the smartphone popularity is increasing, users are often happy to be provided with location-based suggestions. For instance, a user might wish to discover nearby restaurants while traveling, based on his/her latest history of browsing or ratings of other visited restaurants. Furthermore, Chen and Tsai [19] developed a personalized location-based mobile tourism application (PLMTA) for travel planning. The PLMTA utilized hybrid filtering technology by integrating location with users' search intention, to make more efficient customized tourism recommendations. It allows users to more effectively search through travel information for better users' intention understanding. This study showed a great improvement of recommendation accuracy by considering users' contextual data, which helped in generating more relevant and personalized results.

One of the well-known personalized RSs is the study by Bao et al. [20] who proposed a location-based social network RS (LBSN) that helps the users to find interesting restaurants for dining. The recommendation was based on two factors (i) the user's preferences that were learned from his/her location history and (ii) social opinions, which were mined from the location histories of the local experts. The authors found LBSN to improve the recommendation accuracy when both the social-based filtering and geographical location filtering were integrated. The current study intends to extend this concept by using users' geographical location to improve recommendations, with the assumption that when a user travels from Malaysia to Singapore for example, the list of books recommended may differ to include the current trending books in Singapore.

CF does not perform well when there is no sufficient rating data (due to not acquiring enough information from new users) [8], hence Content-based Filtering (CB) technique is often used to overcome the limitations of CF. The item descriptions are used as training data to create a user-specific classification or regression modelling problem. For each user, the training documents correspond to the descriptions of the items he/she has bought or rated [21]. For example, CB was used to improve recommendations based on a series of characteristics from a particular book to further recommend additional books with similar content [22].

CB has gained a lot of attention due to the use of semantic analysis and ontology tools in RSs. Semantic technologies help to improve the recommendation by trying to solve the language ambiguity in textual information. For instance, Garrido & Ilarri [23] proposed a lingual, ontological and semantic enhanced book RS called Topic Map Recommender (TMR) to produce a personalised book recommendation that meets each individual's preferences. The study illustrated the possibility of utilizing CB approach along with textual information such as reviews, descriptions, tags and titles.

Semantic tools help to extract the embedded contextual data from the textual information. For example, in "*I love to read this book during the weekend*", during the weekend is the relative contextual attribute that defines the user's context, hence an RS may recommend similar books ("this book") during the weekend. A similar study was conducted in the medical field, whereby the authors developed MediSem, an RS that semantically examines a patient's information (i.e. background records medical studies), and uses these information along with the user's profile to suggest publications that match the patient's medical records. It showed a great improvement in health

recommendation domain by considering the extracted contextual knowledge about patients [24].

Literature review shows that incorporating users' and products' contextual features improve the RSs, along with an integration of filtering techniques. CAPHyBR aims to enhance book recommendations by manipulating users' features (personality traits, demographic data and geographical location) and product features (review sentiments and purchase intention), using CF, CB and DF. The architecture and implementation of CAPHyBR is presented in the following section.

3 Methodology

This section begins with CAPHyBR's architecture, followed by the recommendation logic and experimental evaluations. The overall system methodology is described in detail, guided by the research goals and objectives.

3.1 System Architecture

Figure 1 depicts the overall CAPHyBR architecture, which can be categorized into three main parts, namely registration, user profiling and recommendation. The flow of the mechanisms involved is described in detail below:

- i. A user provides his/her basic demographic details (i.e. age and gender), and answers the TIPI questionnaire
- ii. CAPHyBR performs user profiling based on the demographic data and personality trait. It monitors the user's behaviour (such as browsing, searching, reviewing) and stores the data in the user profile database. The user is able to search for books based on the title, author name, book genre or any desired keyword. The user is also able to rate the book using the 5-star rating system (1: lowest level; 5: highest level). For example, assume User 1 with an extraversion personality, 25 years old male from Kuala Lumpur, as shown in Figure 2.

The figure shows a more personalized recommendation for User 1 based on his recent rating for the book "Age of Kali Travel" 5 stars (Figure 2: Step 1.2).

When the user is about to purchase a book, he is asked to enter his purchasing reason, which is then stored in the user's profile along with any reviews provided. The user's rating, purchase reason and review will have a strong effect on his next recommendation result. CAPHyBR also

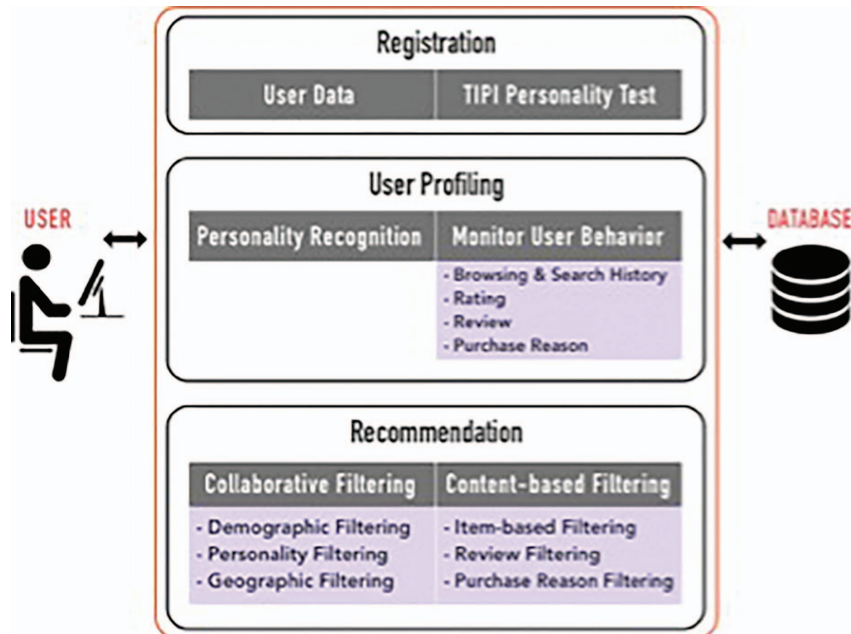


Figure 1 The CAPHyBR architecture.

retrieves the current geographic location data from the device every time the user launches the application. For example, if the user travels from City A to City B, the system will take the last visited place (City B) as the current place and will use it in the recommendation algorithm.

- iii. CAPHyBR generates a personalized book recommendation by combining two main filtering approaches, CF and CB (more details in Section 3.2). It begins with CF encompassing geographic filtering, personality and demographic filtering. Nearby users (within the same city) with the same personality trait and age range are selected, prior to the execution of CB. CB looks for books that have been rated or reviewed by similar users, and a more personalized and relevant books are recommended. For example, assume User 2 has the same characteristics as User 1. CAPHyBR will provide the same recommendation as the list of recommended books is affected by User 1's behaviour (Figure 3).

Assume User 2 now searches the keyword "War" and rates the book "The Strange Death of World War II" as 5 stars. The recommendation list changes according to his latest behaviour (Figure 3: Step 2.3).

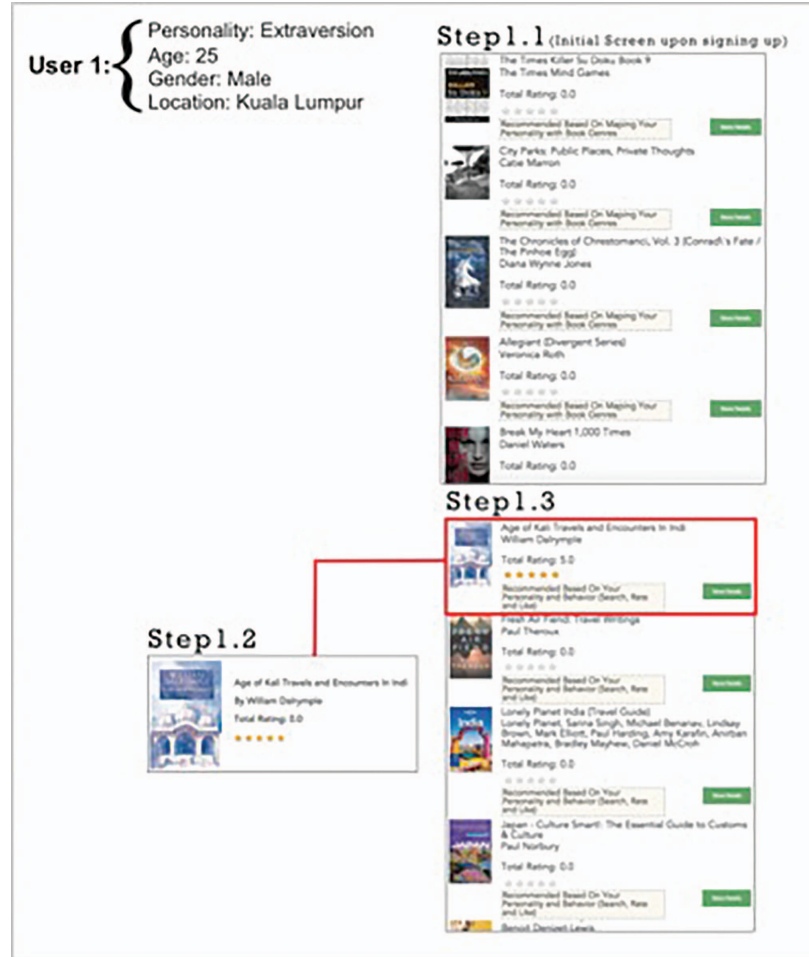


Figure 2 Recommended books for User 1: search history.

3.2 Recommendation Logic

We present the proposed RS's algorithm for CAPHyBR in this section, which produces personalised book recommendations utilising multiple filtering methods, as illustrated in Figure 4.

3.2.1 Collaborative filtering

As Figure 4 shows, there are three filtering methods associated with CF, namely geographical, personality and demographic.

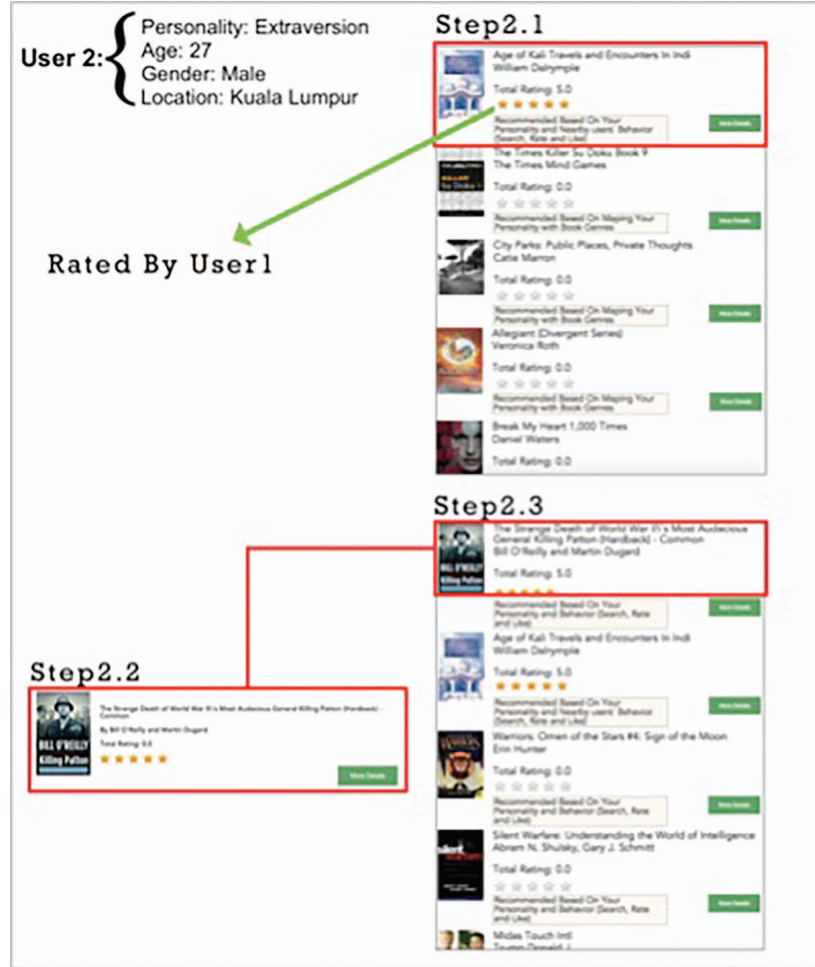


Figure 3 Recommended books for User 2: similar users.

Geographical filtering

CAPHyBR first filters nearby users who are within the same city based on the Haversine formula [25], as depicted in Equation (1), where d is the distance between two locations.

$$d = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right) \quad (1)$$

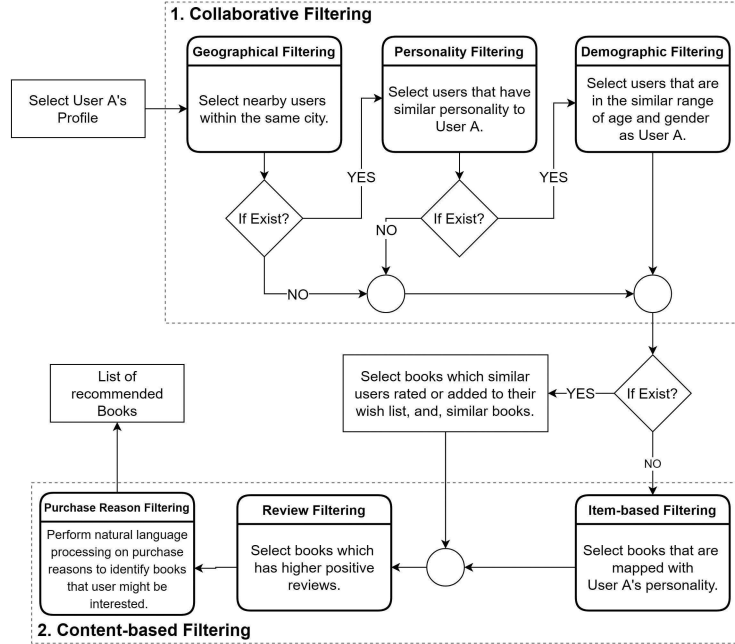


Figure 4 Processing steps of CAPHyBR's recommendation logic.

Every geographical location has a longitude and latitude that is used to specify the location on earth. φ_1 and φ_2 are the latitudes of the first and second location, whereas λ_1 and λ_2 are the longitudes [25]. CAPHyBR obtains the latest geographical location every time a user uses the mobile application to ensure the algorithm uses the latest coordinates.

Personality filtering

When the nearby users are selected, CAPHyBR filters those with similar personality traits. The personalities are then compared with User A's personality. The cosine-based similarity was used to measure the similarity between two users, U_i and U_j , as shown in Equation (2) [26]:

$$\text{Cos}(U_i, U_j) = \frac{U_i \cdot U_j}{\|U_i\| \|U_j\|} \quad (2)$$

By determining the cosine-based similarity, the system is effectively trying to find cosine value of the angle between two users' personalities. The personalities are determined similar when the calculated cosine value is 1.

Demographic filtering

Finally, CAPHyBR filters the list of users using their demographic details (i.e. age and gender). Three age groups were defined in this study: Teen (age <20), Young Adult (age 20–40) and Adult (age >40). Similar to personality filtering, the demographic filtering also used cosine-based similarity approach to measure the users' demographic comparison.

Upon the completion of all three filtering mechanisms, CAPHyBR begins to analyse the user's browsing behaviour, which include visited books, rated books, searched keywords and his wish list, etc. In order to map the book genres with personality traits, CAPHyBR utilized the personality-based user stereotypes suggested by Cantandor et al. [5]. The stereotype distinguishes female and male users, based on 16 genres containing vectors of five real values in the [1, 5] range that correspond to the average scores of the Big Five personality factors. The initial list of recommendations will then go through the second phase, which is CB filtering.

3.2.2 Content-based filtering

This phase contains three more filtering techniques: item-based filtering, review filtering and purchase reason filtering (refer to Figure 4). It is to note that if the CF algorithm is able to generate the list of books based on similar users, then the item-based filtering will be skipped. The filtering will instead continue with the review and purchase reason filtering.

Item-based filtering

The item-based filtering is administered if CAPHyBR is unable to perform CF due to a lack of adequate information about the user. The filtering mechanism maps the user's personality with book genres and filters the books based on similar books that have been rated by the user (Note: books with ratings lower than 3 are not considered). Lastly, it takes the last searched keyword and filters books that contain the keyword in their title or synopsis.

Review filtering

Sentiment analysis which determines the polarity of a sentence as positive, neutral or negative is performed using MeaningCloud, a popular cloud-based software that can be used to extract valuable information from any text source. During this filtering process, CAPHyBR counts the number of positive reviews for each of the books and adds more weight to the book accordingly. However, the negative reviews has the reverse effect. According to Equation (3), the adjusted Bayesian formula is used to calculate the weight

of reviews and ratings (W_{RR}) [6]:

$$W_{RR} = \left(\frac{n}{n + \min_{RR}} \right) \times \bar{R} + \left(\frac{\min_{RR}}{n + \min_{RR}} \right) \times S \quad (3)$$

Where:

- n = number of reviews and ratings for the book
- \min_{RR} = minimum number of reviews and ratings for the book
- \bar{R} = average rating for the book
- S = sum of all reviews of the book

This formula is applied to every book in the list and once the W_{RR} is calculated, it will be added to the book weight. This results in books with higher positive reviews to be placed at the top of the recommendation list.

Purchase reason filtering

This final filtering technique utilizes the user's purchase reason to enhance the recommendation accuracy. CAPHyBR tries to analyse similar nearby users' purchase reason and match them with the user's search history. Thus, more weight will be added to books with purchase reasons related to the user's search profile. In text analysis, large strings of text can be expressed in tokens. These tokens often correspond to words. Figure 5 shows a simple tokenization example for "Motivational and inspiring stories", resulting in four tokens, i.e. "Motivational", "and", "inspiring", "stories". Next, every token is mapped to its root form (i.e. stemming). For example, "Motivational" is mapped to "Motive".

This process is applied to all the purchase reasons of a book, resulting in a list of root words. As Equation (4) shows, CAPHyBR calculates the weight of purchase reason (W_P) by matching the user's search history (H) with the list of root words (L_R) and adds more weight to the book for every match that

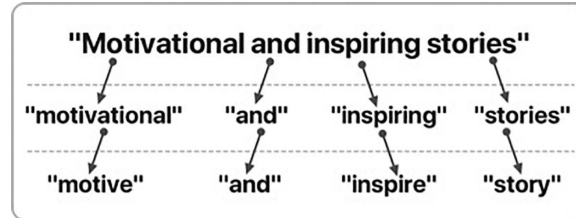


Figure 5 Tokenization and stemming process.

occurs.

$$W_P = |L_R \cap H| \quad (4)$$

The filtering process is now complete, and the final list of book recommendations are provided to the user, sorted based on the books' total weight. The study limits the book recommendations to 15.

3.3 Experimental Evaluation

This section elaborates the experiment and evaluation, including the experimental setup, dataset and evaluation metrics used. We evaluated the effectiveness of CAPHyBR using two forms of evaluations, that is, the overall CAPHyBR's effectiveness, along with the comparisons between each features. To be specific, the comparisons were conducted as below:

- CAPHyBR_D – filtering mechanism using only users' demographic data
- CAPHyBR_DP – the aforementioned mechanism with personality traits included
- CAPHyBR_DPG – the aforementioned mechanism with geographical location included
- CAPHyBR – the aforementioned mechanism with user reviews and purchase reasons included (i.e. complete system)

The second form of evaluation involved a comparison against other existing algorithms, namely Personalized Location-based Mobile Tourism Application (PLMTA), Social Presence on Personalized Recommender System (SPRS), Location-based Recommendation Using Sparse Geo-Social Networking Data (LBSN) and Personalized Medical Reading Recommendation (MediSem). These algorithms were selected as the RSs were similarly based on hybrid algorithms, including a few user and product features. In addition, all these studies were evaluated using the same metrics, hence, a comparison can be performed easily.

3.3.1 Experimental setup

The experiment was implemented from April 3rd 2017 to July 14th 2017, taking place in two computer labs, one in Kuala Lumpur and the other in Malacca (i.e. a different state). A total of 50 participants majoring in Computer Science participated in the experiment, ranging between 19 and 24 years old (Female = 20; Male = 30). CAPHyBR was uploaded to the cloud server to be accessible via the computer's Web Browser software.

A printed user guide was provided, along with a simple demonstration in using CAPHyBR. The participants were given approximately ten minutes to familiarize themselves with the RS before the actual experiment took place. The experiment basically requires the participants to use the RS to search for books based on the keywords and genres provided in the user guide, and to rate the relevance of the books as well. The experiment lasted for approximately 30 minutes.

3.3.2 Evaluation metrics

The effectiveness of CAPHyBR was measured using two popular evaluation metrics: Standardized Root Mean Square Residual (SRMR) and Root Mean Square Error of Approximation (RMSEA). SRMR is the absolute measure of fit, estimated by the square root of the estimated discrepancy due to approximation per degree of freedom [27], as shown in Equation (5).

$$\text{SRMR} = \sqrt{\sum_{i=1}^p \sum_{j=1}^i \left[\frac{s_{ij} - \hat{\sigma}_{ij}}{s_{ii}s_{jj}} \right]^2 / p(p+1)/2} \quad (5)$$

Differences between data (s_{ij}) and model ($\hat{\sigma}_{ij}$) predictions comprise the residuals where p is the total number of observed variables. The average is computed, and the square root taken. SRMR is a badness-of-fit index (i.e. larger values signal worse fit and lower values indicate better performance), which ranges from 0.0 to 1.0. If the SRMR is zero, it means the model predictions match the data perfectly, and a value of 0.08 or less being indicative of an acceptable model [27]. SRMR is enhanced (lowered) when the measurement model is clean (high factor loadings). The index is a good indicator of whether the system captures the data, because it is relatively less sensitive to other issues such as violations of distributional assumptions.

The second metric is RMSEA, as shown in Equation (6):

$$\text{RMSEA} = \sqrt{\max([(\chi^2 / \text{df}) - 1) / (N - 1)], 0)} \quad (6)$$

Where χ^2 is the chi-square value, df is its degrees of freedom and N is the sample size. The RMSEA has a best estimation of zero when the data fits the model. When data is over fitted to the model, $\chi^2 / \text{df} < 1$, is ignored. For a given χ^2 , RMSEA decreases as sample size, N , increases. The RMSEA ranges from 0 to 1, with lower values showing a greater model fit. A value of 0.06 or lower is expressive of acceptable model fit. RMSEA considers two errors: the error of approximation that demonstrates the absence of fit of the

Table 1 List of symbols and equations

Equations	Symbols	Definition
Equation 1	d	Distance
	φ	Latitude
	λ	Longitude
	r	Radius
Equation 2	U_i	User i
Equation 3	W_{RR}	Weight for reviews and ratings
	n	Number of reviews and ratings
	\min_{RR}	Minimum number of reviews and ratings for a book
	\bar{R}	Average rating for a book
Equation 4	W_P	Weight for purchase reason
	S	Sum of all reviews
	H	Search history
	(L_R)	List of root words
	χ^2	Chi-square
	df	Degree of freedom
	N	Sample size

system when the parameter is ideally picked, and the error of estimation that shows the absence of fit of the system to population data [27].

Table 1 shows all the symbols and their respective equations.

4 Results and Discussion

This section begins with the results comparing the various features and filtering mechanisms, followed by the comparison with existing RSs.

4.1 Features and Filtering Mechanisms

Table 2 shows the RMSEA and SRMR values for the varying features and filtering mechanisms.

Table 2 Error rates for the models

Recommendation Systems	Features	RMSEA	SRMR
CAPHyBR_D (A)	Demographic	0.333	0.702
CAPHyBR_DP (B)	Demographic + Personality	0.102	0.422
CAPHyBR_DPG (C)	Demographic + Personality + Location	0.091	0.301
CAPHyBR (D)	Demographic + Personality + Location + Review + Purchase Reason	0.050	0.153

Letters A, B, C and D are mapped to Figure 6.

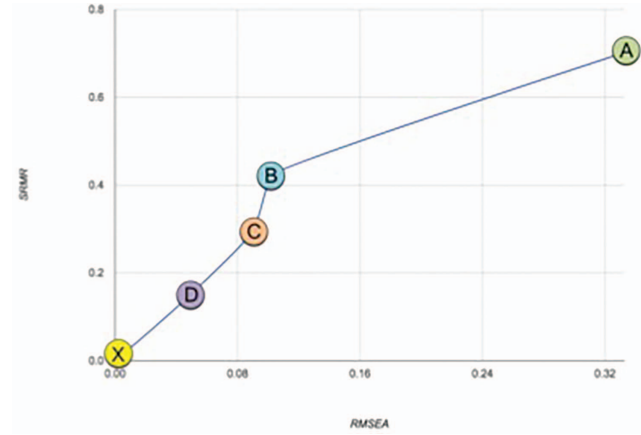


Figure 6 Recommendation performance evaluation.

It can be clearly observed that the overall error rates decreases with the increase of the contextual features. Figure 6 illustrates the results with point X (i.e. origin of coordinates) indicating an absolute accuracy. It can be noted that the RMSEA value decreased from point A to point B by 69%. At point C, the RMSEA evaluation is 0.091, which is 11% lower compared to point B, that is, the recommendation without considering users' geographic locations. Finally, the overall RMSEA value improved when all the filtering mechanisms were integrated (i.e. CAPHyBR), with the error rate dropping to 0.050 (i.e. 44% improvement on the recommendation accuracy).

As the RMSEA value moves closer to 0, it shows a higher accuracy, similar with the SRMR evaluation. It can be observed that the personalized recommendation accuracy is improved when users' demographic data were considered with personality traits, location, review sentiments and purchase reason. RMSEA is considered as "good fit" when its value is equal or less than 0.05, hence indicating CAPHyBR to be a good fit model.

Prior studies have showed the effectiveness of contextual-features when tested in silo [1, 4, 5, 8, 21]. For instance, Xin et al. [28] achieved a higher book recommendation accuracy when they took users' personality traits into consideration. In addition, Guan et al. [29] and Gil et al. [30] who improved book recommendations based on users' personality traits echoed similar findings. Furthermore, the product textual features are important to be considered during recommendation process and by integrating them with users' contextual data, a higher recommendation accuracy can be achieved. This was clearly shown in this study based on the highest accuracy produced

Table 3 Comparison results between various recommendation systems

Recommendation Systems	RMSEA
Context-Aware Personalized Hybrid Book Recommender (CAPHyBR)	0.050
Personalized Location-based Mobile Tourism Application (PLMTA)	0.063
Social Presence on Personalized Recommender System (SPRS)	0.070
Location-based Recommendation Using Sparse Geo-Social Networking Data (LBSN)	0.055
Personalized Medical Reading Recommendation – Deep Semantic Approach (MediSem)	0.060

for CAPHyBR whereby both user and product features were integrated to improve recommendation.

4.2 CAPHyBR Comparison with Existing RSs

Table 3 below shows the comparison results between CAPHyBR and the existing RSs, namely, PLMTA, SPRS, LBSN and MediSem.

Looking at Table 3, it can be observed that CAPHyBR outperformed the rest of the RSs as it produced the lowest error rate, with the closest contender being LBSN and MediSem. LBSN works by integrating social-based filtering with geographical location filtering to achieve better accuracy recommendation. LBSN with RMSEA value of 0.055, showed that incorporating multiple related filtering methods can improve the RS's accuracy. Location-based filtering is one of the key factors in tourism and social applications [19]. However, it has been shown in other studies that geographical location can be used in other domains to improve recommendation accuracy as well [20]. This filtering method had a great impact on CAPHyBR's performance and as it was used to group the users in correlation of other CF filtering methods (i.e. personality and demographic filtering).

Erekhinskaya et al. [24] showed a great progress in their work (MediSem) with an RMSEA value of 0.060 which yields a better model fit compared to other medical RSs. They performed the recommendations based on the extracted knowledge that was gathered semantically from a patient's profile. It analysed the textual records and medical articles, then performed deep semantic extraction. Semantic text analyses approaches need to be used to extract meaningful and valuable insights about a product's textual content for better understanding users' context and interests. Previous studies have showed that a context-aware RS makes the recommendations more aligned with the intended context of user. Therefore, it can be observed that when

CAPHyBR applies sentiment analysis on the user reviews to identify most relative books to recommend, it results in an improvement.

The development of more complete model of users' interests and behaviours opens up an opportunity for the development of more sophisticated techniques for personalisation that more accurately capture users' context in the real world. Aside from the mentioned algorithms' results, CAPHyBR showed that when several effective filtering techniques were integrated, recommendations can be greatly improved, and thus resulting in a much improved user satisfaction as well.

5 Conclusion, Limitation and Future Work

Incorporating contextual information in RS is an effective approach to create a more personalized, accurate and relevant recommendations. Therefore, this study integrated several users' characteristics, namely their personality traits, demographic details and current location, together with product features (i.e. review sentiments and purchase reasons) to improve book recommendations. Results indicate that these contextual-features can be used effectively to improve the recommendation accuracy, with overall error rates of 0.050 and 0.150 for RMSEA and SRMR, respectively.

The study, however, has its limitations. First, the experiments conducted involved users who are technologically savvy (i.e. Computer Science background), therefore they had the technical skill to easily learn and use CAPHyBR to locate the books that are relevant to them. It would be interesting to evaluate CAPHyBR with users with no or low technology skill to determine the effectiveness of CAPHyBR in recommending relevant books. Future studies therefore, could replicate the current study and expand it to include users from various background and digital skill.

Second, it was rather challenging to test CAPHyBR in different countries (i.e. Thailand and Singapore) for the geographical location feature, due to resource constraints. A best scenario is in which evaluations can be conducted with users traveling between more geographically distanced states, or neighbouring countries. This will then provide a clearer impact of location in recommending books to the users.

References

- [1] Bhosale, S., Nimse, P., Wadgaonkar, S., & Yeole, A. (2017). SuggestA-Book: A Book Recommender Engine with Personality based Mapping. *International Journal of Computer Applications*, 159(9), 1–4.
- [2] Kim, E., Kim, S.-Y., Kim, G.-A., Kim, M., Rho, S., Man, K. L., & Chong, W. K. (2015). TiMers: Time-based Music Recommendation System based on Social Network Services Analysis. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (Vol. 2).
- [3] Winoto, P., & Tang, T. Y. (2010). The role of user mood in movie recommendations. *Expert Systems with Applications*, 37(8), 6086–6092.
- [4] Braunhofer, M., Elahi, M., & Ricci, F. (2014). Usability Assessment of a Context-Aware and Personality-Based Mobile Recommender System, In: Hepp M., Hoffner Y. (eds.) E-Commerce and Web Technologies. EC-Web 2014. *Lecture Notes in Business Information Processing*, 188. Springer, Cham.
- [5] Cantador, I., Fernández-tobías, I., & Bellogín, A. (2013). Relating Personality Types with User Preferences in Multiple Entertainment Domains. *Proceedings of the 1st Workshop on Emotions and Personality in Personalized Services (EMPIRE)*, (pp. 1–16).
- [6] Chen, Q., Zheng, S., Chen, H., & Liu, W. (2017). Research on Recommendation Mode of WeChat Official Account Platform Based on Hybrid Recommendation Algorithm. *DEStech Transactions on Environment, Energy and Earth Sciences*, 57–61.
- [7] Nirwan, H., Verma, O. P., & Kanojia, A. (2016). Personalized hybrid book recommender system using neural network. In *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on* (pp. 1281–1288).
- [8] Zhang, Y., Chen, M., Huang, D., Wu, D., & Li, Y. (2017). iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. *Future Generation Computer Systems*, 66, 30–35.
- [9] Arabi, H., & Balakrishnan, V. (2019). Personalized Hybrid Book Recommender. *International Journal of Information Systems in the Service Sector*, 11 (3), 70–97.
- [10] Lu, C.-C., & Tseng, V. S. (2009). A novel method for personalized music recommendation. *Expert Systems with Applications*, 36(6), 10035–10044.

- [11] McCrae, R. R., & Costa, P. T. (1996). Toward a New Generation of Personality Theories: Theoretical Contexts for the Five-Factor Model. [In] Wiggins JS (Ed.): *The Five-Factor Model of Personality: Theoretical Perspectives*. Guilford, New York.
- [12] Chiorri, C., Bracco, F., Piccinno, T., Modafferi, C., & Battini, V. (2015). Psychometric Properties of a Revised Version of the Ten item Personality Inventory. *European Journal of Psychological Assessment*, 31(2), 109–119.
- [13] Yang, B., Lei, Y., Liu, J., & Li, W. (2016). Social collaborative filtering by trust. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(8), 1633–1647.
- [14] Luo, C., Pang, W., Wang, Z., & Lin, C. (2014). Hete-cf: Social-based collaborative filtering recommendation using heterogeneous relations. In *2014 IEEE International Conference on Data Mining* (pp. 917–922).
- [15] Guo, X., Feng, L., Liu, Y., & Han, X. (2016). Collaborative filtering model of book recommendation system. *International Journal of Advanced Media and Communication*, 6(2–4), 283–294.
- [16] Choi, J., Lee, H. J., & Kim, Y. C. (2011). The Influence of Social Presence on Customer Intention to Reuse Online Recommender Systems: The Roles of Personalization and Product Type, *International Journal of Electronic Commerce*, 16(1), 129–154.
- [17] Zhao, W. X., Li, S., He, Y., Wang, L., Wen, J.-R., & Li, X. (2016). Exploring demographic information in social media for product recommendation. *Knowledge and Information Systems*, 49(1), 61–89.
- [18] Rich, E. (1979). User modeling via stereotypes. *Cognitive Science*, 3(4), 329–354.
- [19] Chen, C.-C., & Tsai, J.-L. (2017). Determinants of behavioral intention to use the Personalized Location-based Mobile Tourism Application: An empirical study by integrating TAM with ISSM. *Future Generation Computer Systems*, 9, 628–638.
- [20] Bao, J., Zheng, Y., & Mokbel, M. F. (2012). Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems* (pp. 199–208).
- [21] Lu, Z., Dou, Z., Lian, J., Xie, X., & Yang, Q. (2015). Content-Based Collaborative Filtering for News Topic Recommendation. In *AAAI* (pp. 217–223).
- [22] Mathew, P., Kuriakose, B., & Hegde, V. (2016). Book Recommendation System through content based and collaborative filtering method.

- In *Data Mining and Advanced Computing (SAPIENCE)*, *International Conference on* (pp. 47–52).
- [23] Garrido, A. L., & Ilarri, S. (2014). TMR: a semantic recommender system using topic maps on the items' descriptions. In *European Semantic Web Conference* (pp. 213–217).
 - [24] Erekhinskaya, T., Balakrishna, M., Tatu, M., & Moldovan, D. (2016). Personalized Medical Reading Recommendation: Deep Semantic Approach. In *International Conference on Database Systems for Advanced Applications* (pp. 89–97).
 - [25] Ratsameethammawong, P., & Kasemsan, K. (2010). Mobile Phone Location Tracking by the Combination of GPS, Wi-Fi and Cell Location Technology. *Communications of the IBIMA. 2010*, Article ID 566928.
 - [26] Kumar, A., Gupta, S., Singh, S. K., & Shukla, K. K. (2015). Comparison of various metrics used in collaborative filtering for recommendation system. In *Contemporary Computing (IC3), 2015 Eighth International Conference on* (pp. 150–154).
 - [27] Kelley, K. & Lai, K. (2011) Accuracy in Parameter Estimation for the Root Mean Square Error of Approximation: Sample Size Planning for Narrow Confidence Intervals, *Multivariate Behavioral Research*, 46, 1–32.
 - [28] Xin, L., Haihong, E., Junjie, T., Meina, S., & Yi, L. (2014). Enhancing Book Recommendation with Side Information. In *Service Sciences (ICSS), 2014 International Conference on* (pp. 142–146).
 - [29] Guan, C., Guan, C., Qin, S., Qin, S., Ling, W., Ling, W., . . . Ding, G. (2016). Apparel recommendation system evolution: an empirical review. *International Journal of Clothing Science and Technology*, 28(6), 854–879.
 - [30] Gil, J.-M., Lim, J., & Seo, D.-M. (2016). Design and Implementation of MapReduce-Based Book Recommendation System by Analysis of Large-Scale Book-Rental Data. In J. J. (Jong H. Park, H. Jin, Y.-S. Jeong, & M. K. Khan (Eds.), *Advanced Multimedia and Ubiquitous Engineering: FutureTech {&} MUE* (pp. 713–719). Singapore: Springer Singapore.

Biographies



Hossein Arabi received his Ph.D in information system from University of Malaya in 2018. He is currently working as a solution architect which is leading the design, development, implementation and testing of IT solutions. Over the years, he has been involved in development of variety systems, from monolithic architecture to modular-based and now on micro-service architecture.



Vimala Balakrishnan is an Associate Professor and a Fulbright Research Scholar affiliated with the Faculty of Computer Science and Information Technology, University of Malaya since 2010. She obtained her PhD. in the field of Ergonomics from Multimedia University, whereas her Masters and Bachelor degrees were from University of Science Malaysia. Dr Balakrishnan's main research interests are in data analytics and sentiment analysis, particularly related to social media. Her research domains include healthcare, education and social issues such as cyberbullying. She has published approximately 60 articles in top indexed journals and 44 conference proceedings, has four patents and eight copy-rights.



Nor Liyana Mohd Shuib obtained her Master of Information System (Data Mining) from Universiti Kebangsaan Malaysia in 2005 and a Ph.D. from the University of Malaya, Malaysia in 2013 respectively. She is a Senior Lecturer at the Department of Information Systems, Faculty of Computer Science & Information Technology, University of Malaya, Malaysia. She has published a number of journal papers and proceedings locally and internationally. Her research interests include personalization, e-learning, recommender system, data science, data mining, artificial intelligence application, and educational technology. She has won more than 20 awards from reputable innovation competition internationally. She is also a Senior Member of IEEE computing society, an active blogger and presently, the principal investigator of multiple research grant in the Faculty.

