

# Book Recommendation System

**Kieu.M.Phuoc, Phan.H.B.Khang, Phan.T.Nhan, Huynh.V.Tin**

Faculty of Information Science and Engineering, University of Information Technology,  
Ho Chi Minh City, Vietnam

Vietnam National University, Ho Chi Minh City, Vietnam

{20521774, 20520258, 20520570} gm.uit.edu.vn  
tinhv@uit.edu.vn

## Abstract

Sách là nguồn tri thức quý báu của nhân loại. Vì thế việc tìm được một cuốn sách phù hợp với nhu cầu của mỗi người là rất quan trọng. Dựa vào nhu cầu trên, một recommender system được phát triển nhằm mục đích khuyến nghị những cuốn sách thích hợp đến cho người dùng. Hệ khuyến nghị là hệ thống phân tích khôi dữ liệu người dùng và đưa ra dự đoán, gợi ý đề xuất được cho là phù hợp với sở thích của người dùng tại thời điểm bất kỳ trên các ứng dụng và nền tảng trực tuyến giúp tiết kiệm thời gian tiềm kiếm, truy cập nội dung dễ dàng đồng thời giúp nâng cao trải nghiệm khách hàng. Nhóm đã sử dụng các phương pháp tiếp cận là lọc dựa trên cộng tác - Collaborative Filtering (CF), lọc dựa trên nội dung - content-based filtering (CBF), lọc theo tri thức - Knowledge-based Filtering, lọc theo top thịnh hành - Demographic Filtering và phương pháp kết hợp - hybrid (CBF+CF).

## 1 Giới thiệu

Trong thời đại hiện nay, khi công nghệ ngày càng phát triển, hệ khuyến nghị (Recommender Systems) đóng vai trò quan trọng trong việc tối ưu hóa trải nghiệm đọc sách của người đọc. Hệ thống khuyến nghị được ứng dụng rất nhiều trong phân tích dữ liệu người dùng và dự đoán, gợi ý, đưa ra nội dung đề xuất cho người dùng. Hệ thống này không chỉ giúp chúng ta khám phá những tác phẩm mới mà còn tối ưu hóa sở thích và nhu cầu cá nhân của từng độc giả, từ đó mở rộng và đa dạng hóa kiến thức, văn hóa đọc trong xã hội. Một số hệ thống khuyến nghị tiêu biểu phải kể đến là hệ thống của Amazon, Netflix và Youtube. Ở Việt Nam ngày nay hệ khuyến nghị cũng được áp dụng nhiều vào lĩnh vực thương mại điện tử như Shopee, Tiki, Lazada. Như vậy, hệ thống khuyến nghị hỗ trợ người dùng ra quyết định và gợi ý sản phẩm đến người dùng, giúp tiết kiệm thời gian, tăng tốc độ tìm kiếm và giúp người dùng truy cập nội dung họ quan tâm một cách dễ dàng hơn.

Một trong những lợi ích quan trọng nhất của hệ khuyến nghị sách chính là khả năng tiếp cận những cuốn sách phù hợp với độc giả. Thay vì phải dành thời gian tìm kiếm thông tin, hệ thống này sẽ tự động đề xuất những tác phẩm mà người đọc có thể quan tâm, dựa trên lịch sử đọc sách và sở thích cá nhân. Điều này không chỉ giúp tiết kiệm thời gian mà còn tạo ra trải nghiệm đọc sách cá nhân hóa, làm tăng độ hứng thú và tiện ích cho người đọc.

Có hai loại hệ thống gợi ý truyền thống chính: khuyến nghị lọc theo cộng tác (CF) và khuyến nghị dựa trên nội dung (CB). Khi hệ thống khuyến nghị lọc theo cộng tác dự đoán liệu người dùng có thích một cuốn sách hay không, nó sẽ dựa vào xếp hạng của tất cả người dùng đang hoạt động trong hệ thống. Khuyến nghị lọc theo cộng tác phụ thuộc vào ma trận xếp hạng có thể lớn và thưa thớt, và điều đó có thể dẫn đến những khuyến nghị kém hơn. Sự thưa thớt xảy ra khi các mục hoặc người dùng không có đủ xếp hạng. Có thể nhận thấy trong hồ sơ thư viện rằng nhiều cuốn sách không bao giờ được mượn.

Mặt khác, khuyến nghị dựa trên nội dung không yêu cầu ma trận đánh giá; điều đó loại bỏ tất cả các vấn đề đã nói ở trên với khuyến nghị lọc theo cộng tác. Tuy nhiên, khuyến nghị dựa trên nội dung khó có thể giới thiệu những mặt hàng có nội dung chưa đầy đủ. Điểm yếu khác là chuyên môn hóa quá mức, xảy ra khi hệ thống đưa ra các khuyến nghị không đa dạng. Hệ khuyến nghị lai có thể kết hợp khuyến nghị lọc theo cộng tác và khuyến nghị dựa trên nội dung để giải quyết các vấn đề như vậy và nâng cao khuyến nghị sự chính xác.

Báo cáo này bao gồm các nội dung như sau: các công trình liên quan ở mục 2, giới thiệu dữ liệu sử dụng ở mục 3, tiếp đến là phương pháp tiếp cận ở mục 4, phân tích kết quả mục 5 và cuối cùng là kết luận và hướng phát triển ở mục 6.

## 2 Các công trình liên quan

Chủ đề nghiên cứu này là đề tài phổ biến trong những năm gần đây. Đặc biệt là trong thời đại bùng nổ về dữ liệu người dùng và các trang web mua sắm trực tuyến. Việc khuyến nghị sản phẩm phù hợp với người dùng đã dần trở thành một trong những yếu tố quan trọng của việc customer service. Nhóm nghiên cứu đã tham khảo và sử dụng nhiều công trình có liên quan. Trong đó có các công trình như:

- Revisiting-the-Tag-Relevance-Prediction-Problem, nhóm nghiên cứu sử dụng bộ dữ liệu giống với công trình này, đồng thời tham khảo qua thuật toán khuyến nghị dựa trên điểm số của các tag sách.(Kotkov et al., 2022)
- Word2vec, đây là mô hình được đào tạo trước của Google, được nhóm sử dụng trong thuật toán khuyến nghị dựa trên nội dung.(Mikolov et al., 2013)
- Apache Spark, đây là framework được dùng để xử lý big data. Nhóm nghiên cứu đã sử dụng Spark cho thuật toán ALS trong khuyến nghị dựa trên lọc cộng tác do bộ dữ liệu quá lớn không thể sử dụng các hàm thông thường cho thuật toán này. (Zaharia et al., 2010)
- Goodread.com<sup>1</sup>, đây là trang web đánh giá sách cho cộng đồng. Nhóm nghiên cứu đã dựa trên trang web này để xây dựng giao diện người dùng để tương tác cho nghiên cứu này.

## 3 Dữ liệu

Trong đồ án này, nhóm nghiên cứu sử dụng bộ dữ liệu **Book Genome Dataset**<sup>2</sup> từ GroupLens. Book Genome Dataset bao gồm một loạt các thông tin về các tác phẩm văn học, từ thông tin cơ bản như tiêu đề, tác giả, ngày xuất bản đến các yếu tố phức tạp như thể loại, chủ đề, ngôn ngữ sử dụng và mối quan hệ giữa các tác phẩm. Trong đồ án này sử dụng 4 file: metadata.json, ratings.json, tags.json, tag\_count.json. Bộ dữ liệu gồm 5152656 lượt xếp hạng (rating) của 350332 người dùng (user) và 239252 lượt gắn thẻ(tag) trên 9374 cuốn sách (book).

- metadata.json: gồm 9374 dòng dữ liệu tương ứng với 9374 quyển sách với các thông tin sách được lưu trữ có 8 trường:

<sup>1</sup><https://www.goodreads.com/>

<sup>2</sup><https://grouplens.org/datasets/book-genome/>

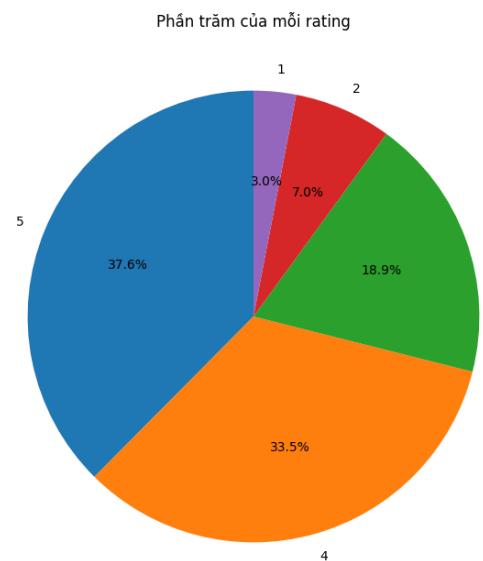


Figure 1: Phân phối của rating

- item\_id: id của sách.
- url: đường dẫn đến trang sách.
- title: tên sách.
- authors: tác giả sách.
- lang: ngôn ngữ của sách.
- img: đường dẫn tới ảnh bìa sách.
- year: năm xuất bản của sách.
- description: mô tả sách.
- ratings.json: gồm 5152656 dòng dữ liệu , là đánh giá của 350332 người dùng tới 9374 cuốn sách. Thông tin lưu trữ có 3 trường: mã sách, mã người dùng thực hiện đánh giá và giá trị đánh giá của người dùng.
- tags.json: gồm 727 dòng dữ liệu, là các thẻ tag của sách. thông tin lưu trữ có 2 trường: mã thẻ tag và tên thẻ tag.
- tag\_count.json: gồm 239252 dòng dữ liệu, là các lượt gắn thẻ tag cho các cuốn sách. thông tin lưu trữ có 2 trường: mã thẻ tag và mã sách.

Hình 1 mô tả phân phối xếp hạng của người dùng theo bậc từ 1 đến 5. Ta thấy dữ liệu phân bố ở các mức rating không đồng đều, tập trung nhiều nhất ở 4.0 và 5.0.

Hình 2 và hình 3 là phân phối của user và sách trong file rating.json.

Hình 4 và Hình 5 là phân phối của sách và các thẻ tag trong file tag\_count.json.

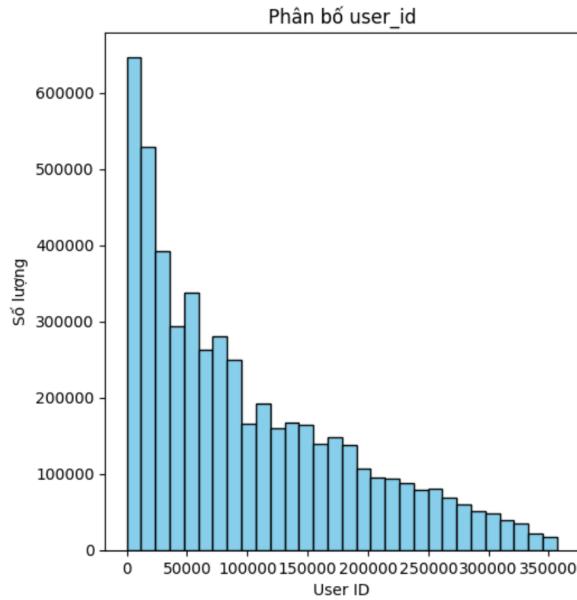


Figure 2: Phân bố của user

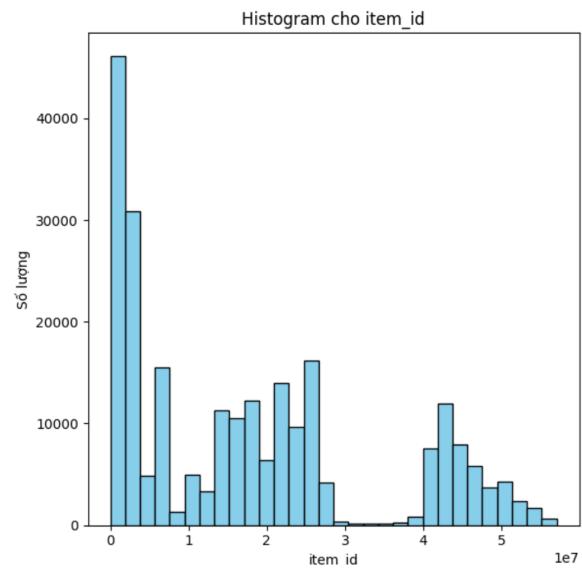


Figure 4: Phân bố của item

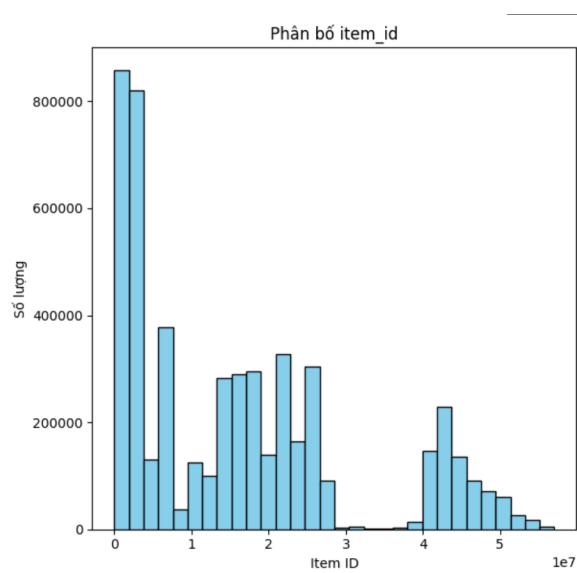


Figure 3: Phân bố của item trên file rating

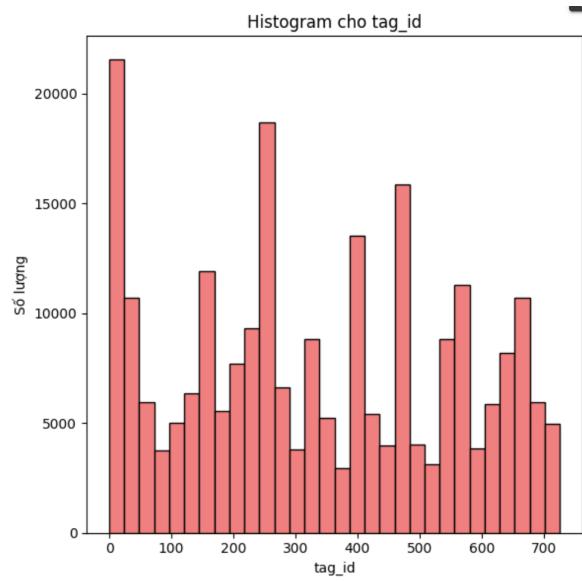


Figure 5: Phân bố của các thẻ tag

## 4 Thí nghiệm

### 4.1 Phương pháp lọc cộng tác

Phương pháp khuyến nghị dựa trên lọc cộng tác là một trong những phương pháp phổ biến trong lĩnh vực khuyến nghị thông tin và tiếp thị. Nó dựa trên nguyên tắc rằng thông tin từ những người dùng tương tự có thể được sử dụng để tạo ra các khuyến nghị cá nhân hóa cho các người dùng khác. Thay vì xem xét chỉ các thuộc tính của người dùng hoặc sản phẩm, phương pháp này tập trung vào việc khai thác thông tin từ sự tương tác giữa người dùng và sản phẩm để xác định mức độ quan tâm của người dùng đối với các sản phẩm khác.

Có nhiều phương pháp lọc cộng tác khác nhau, như Lọc cộng tác dựa trên hàng xóm gần nhất (Nearest Neighbor Collaborative Filtering), Lọc cộng tác dựa trên phân tích ma trận (Matrix Factorization Collaborative Filtering), Lọc cộng tác dựa trên mạng xã hội (Social Collaborative Filtering), và nhiều hơn nữa. Mỗi phương pháp có ưu điểm và hạn chế riêng, và sự lựa chọn phương pháp phù hợp phụ thuộc vào bối cảnh và yêu cầu cụ thể của hệ thống khuyến nghị.

Trong thí nghiệm này, nhóm nghiên cứu sử dụng hai thuật toán đó: Lọc cộng tác dựa trên phân tích ma trận (Matrix Factorization) và lọc cộng tác dựa trên k hàng xóm gần nhất (k-Nearest Neighbor).

Đầu tiên là thuật toán ALS (Alternating Least Squares), đây là một phương pháp quan trọng trong lọc cộng tác dựa trên ma trận. Nó được sử dụng rộng rãi để xây dựng các hệ thống khuyến nghị cá nhân hóa và dựa trên sự tương tác giữa người dùng và sản phẩm để tạo ra các khuyến nghị chất lượng cao.

Thuật toán ALS hoạt động bằng cách tối ưu hóa một hàm mất mát dựa trên ma trận đánh giá. Ý tưởng chính của thuật toán là xác định hai ma trận, một ma trận người dùng và một ma trận sản phẩm, sao cho tích của hai ma trận này xấp xỉ với ma trận đánh giá đã biết nhất có thể. Quá trình này được thực hiện thông qua việc lặp đi lặp lại hai bước: cập nhật ma trận người dùng và cập nhật ma trận sản phẩm (xem hình 3).

Tiếp theo là thuật toán lọc cộng tác dựa trên k hàng xóm gần nhất (k-Nearest Neighbor), đây là một thuật toán phân loại và dự đoán đơn giản và phổ biến trong lĩnh vực học máy. Nó dựa trên nguyên tắc rằng các điểm dữ liệu tương đồng trong không gian đặc trưng sẽ có cùng nhãn hoặc giá trị dự đoán. Thuật toán KNN được sử dụng rộng rãi trong các bài toán phân loại, dự đoán và gợi ý.

Cách hoạt động của thuật toán KNN rất đơn giản. Khi có một điểm dữ liệu mới cần được phân loại hoặc dự đoán, thuật toán KNN tìm kiếm trong tập dữ liệu huấn luyện các điểm gần nhất với điểm đó (k điểm gần nhất, trong đó k là một số nguyên dương đã cho). Sau đó, thuật toán sử dụng đa số phiếu bầu (voting) hoặc trung bình các giá trị của các điểm láng giềng để xác định nhãn hoặc giá trị dự đoán cho điểm dữ liệu mới. Quá trình này dựa trên giả định rằng các điểm gần nhau trong không gian đặc trưng có khả năng thuộc cùng một lớp hoặc có giá trị dự đoán tương tự (xem hình 4).

Đối với cách tính giá trị tương đồng, nhóm nghiên cứu lựa chọn độ đo cosine làm thang đo để tính giá trị tương đồng giữa các người dùng (với 0 là thấp nhất và 1 là cao nhất).

### 4.2 Phương pháp lọc dựa trên tri thức

Phương pháp lọc dựa trên tri thức sử dụng kiến thức và thông tin từ các nguồn bên ngoài để hỗ trợ quá trình phân loại, dự đoán hoặc lọc dữ liệu. Kiến thức này có thể được thu thập từ các chuyên gia trong lĩnh vực tương ứng, các tài liệu tham khảo hoặc các nguồn dữ liệu khác. Phương pháp này có thể sử dụng các thuật toán máy học để học từ tri thức và áp dụng nó vào quá trình phân loại hoặc dự đoán.

Nhóm nghiên cứu sử dụng lịch sử đã đọc các cuốn sách của người dùng để đưa ra dự đoán các cuốn sách người đọc có thể sẽ thích. Các bước bao gồm:

- **Bước 1:** Lấy ra danh sách các cuốn sách mà người dùng đã đọc
- **Bước 2:** Lọc ra các tag sách xuất hiện nhiều nhất trong số các cuốn sách người dùng đã đọc
- **Bước 3:** Lọc ra các cuốn sách chứa các tag đó mà người dùng chưa đọc
- **Bước 4:** Xếp hạng các cuốn sách đó theo rating trung bình của mỗi cuốn sách
- **Bước 5:** Lấy ra k cuốn sách có rating cao nhất và khuyến nghị cho người dùng

### 4.3 Phương pháp lọc nhân khẩu học

Lọc nhân khẩu học (Demographic Filtering) đưa ra đề xuất tổng quát cho người dùng dựa trên mức độ phổ biến của sách. Phương pháp phù hợp cho tập người dùng mới cho có nhiều thông tin về họ. Đây là cách tiếp cận đơn giản nhất để khuyến nghị

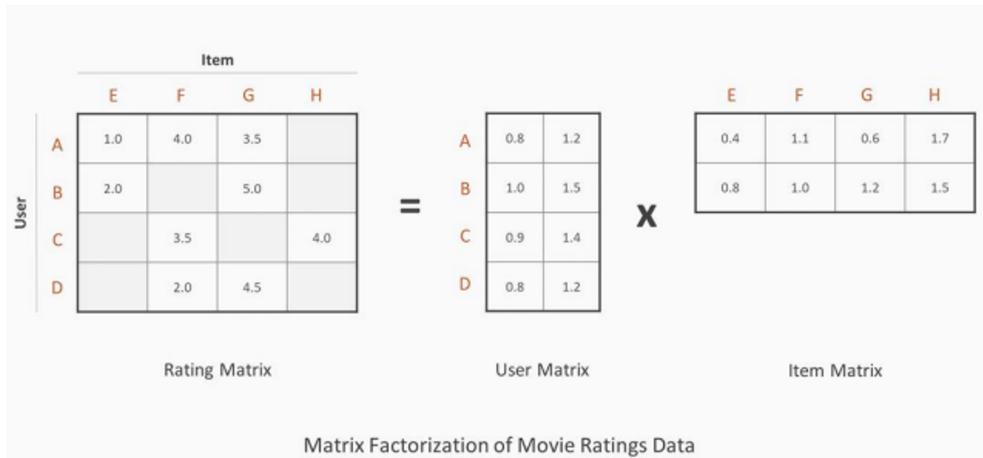


Figure 6: Phân tích ma trận đánh giá

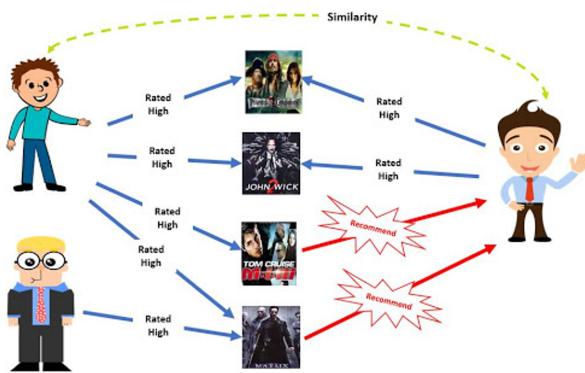


Figure 7: Thuật toán KNN

những sách có đặc điểm nhân khẩu học tương tự đến người dùng khác nhau.

$$\text{Weighted Rating (WR)} = \left( \frac{v}{v+m} \cdot R \right) + \left( \frac{m}{v+m} \cdot C \right)$$

Figure 8: Weight rating (WR)

Phương pháp sẽ đề xuất top k sách có xếp hạng có trọng số (Weighted rating) cao. Quá trình tính toán và đề xuất sách cho người dùng thực thi quá 4 bước: Xác định rating trung bình (R) và số lượng rating (v) cho mỗi quyển sách ở bước 1. Xác định số lượng rating tối thiểu (m) phù hợp và rating trung bình tổng thể (C) ở bước 2. Xác định weighted rating (WR) cho mỗi quyển sách ở bước 3. Và, Xác định top k quyển sách ở bước 4.

- Bước 1:** Dựa vào bảng ratings, tính rating trung bình (R) và số lượng rating (v) cho mỗi quyển sách.

- Bước 2:** Xác định số lượng rating tối thiểu (m) phù hợp và rating trung bình tổng thể (C). Hiện tại ở bộ dữ liệu của chúng tôi, rating trung bình tổng thể (C) là 3.9 trên thang điểm từ 1 đến 5, số lượng rating tối thiểu (m) được xác định bằng cách sử dụng phân vị thứ 90 làm điểm giới hạn của mình. Nói cách khác, để một quyển sách lọt vào bảng xếp hạng, sách đó phải có nhiều lượt bình chọn hơn ít nhất 90% số sách trong danh sách (m hiện tại là 1025.7).

- Bước 3:** Với mỗi quyển sách, tính chỉ số để xếp hạng cho chúng theo công thức weighted rating (WR) (xem hình 8). Trong đó, v là số lượt rating của mỗi quyển sách, m là số lượt rating tối thiểu phù hợp, R là rating trung bình của mỗi sách, C là rating trung bình.

- Bước 4:** Xếp hạng các quyển sách theo weighted rating (wr) và lấy top k sách có điểm số cao nhất để khuyến nghị cho người dùng.

#### 4.4 Phương pháp lọc dựa trên nội dung

Lọc dựa trên nội dung sử dụng các tính năng của mục để đề xuất các mục khác tương tự với những gì người dùng thích, dựa trên hành động trước đó của họ hoặc phản hồi rõ ràng. (Xem hình 9.)

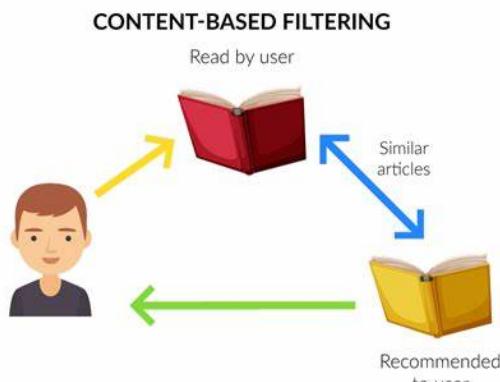


Figure 9: Phương pháp lọc dựa trên nội dung

Dựa trên lịch sử người dùng, chúng tôi đề xuất các quyển sách có nội dung liên quan với nội dung mà người dùng quan tâm. Quy trình thực hiện gồm 4 bước (xem hình 10) là tiền xử lý dữ liệu (Data Preprocessing) trước khi tính toán, chuyển hóa vector (Vectorize) các văn bản đầu vào, tính toán độ tương đồng (Calculate cosine similarity) giữa các quyển sách với nội dung người dùng, xếp hạng và đề xuất (Ranking and recommend) các sách liên quan nhất.

- **Data Preprocessing:** Các dòng dữ liệu đầu vào cần làm sạch trước khi thực hiện tính toán. Quá trình làm sạch bao gồm: Chúng tôi dùng biểu thức chính quy để loại bỏ dấu câu, ký tự đặt biệt. Các stopword được loại bỏ trong quá trình tính toán. Có 790 từ stopword được liệt kê trong danh sách. Tiếp theo văn bản chuyển về chữ thường. Để xác định đúng nghĩa trong quá trình vector hóa, văn bản được tách từ bằng bộ word tokenize của Natural Language Toolkit (NLTK).
- **Vectorize:** Để xác định độ tương đồng giữa các văn bản, dữ liệu đầu vào được biểu diễn thành vector bằng bộ Word2Vec từ mô hình được đào tạo trước của Google (**GoogleNews-vectors**)<sup>3</sup>. Mô hình được tải lên bằng thư viện Gensim.

<sup>3</sup><https://code.google.com/archive/p/word2vec/>

- **Calculate cosine similarity:** Độ tương đồng được tính bằng cosine. Thực hiện tính toán độ tương đồng giữa nội dung của người dùng và tất cả các sách còn lại mà người dùng chưa xem. Kết quả thu được là danh sách độ tương đồng.

- **Ranking and recommend:** Ở bước này, chúng tôi sẽ sắp xếp lại danh sách độ tương đồng và lấy ra k sách có độ tương đồng cao nhất để xuất đến cho người dùng.

#### 4.5 Phương pháp lai

Nhằm tăng hiệu suất của hệ thống, chúng tôi kết hợp các ưu điểm của nhiều phương pháp lai với nhau với mong muốn mang lại hiệu quả đề xuất tốt hơn cho người dùng. Cụ thể, phương pháp lai (Hybrid Filtering) là sự kết hợp giữa phương pháp lọc cộng tác và lọc nội dung. Quy trình thực hiện của phương pháp lai gồm 2 bước như hình 11.

- **Bước 1:** Tìm ra  $n*k$  sách liên quan nhất với nội dung người dùng quan tâm dựa vào phương pháp lọc dựa trên nội dung.
- **Bước 2:** Từ đầu ra của bước 1, dự đoán rating của người dùng với các quyển sách có trong danh sách liên quan. Sau đó, chọn ra top k sách có dự đoán rating cao nhất trong danh sách liên quan để đề xuất cho người dùng.

### 5 Phân tích kết quả

Đối với phương pháp lọc nội dung và phương pháp lai, chúng tôi đánh giá dựa trên hai độ đo là Mean Reciprocal Rank (MRR) và Normalized Discounted Cumulative Gain (NDCG). Việc đánh giá thực hiện trên 790 người dùng có số lượt rating từ 300 và trên 100 sách đề xuất cho mỗi người dùng.

| Phương pháp   | MRR          | NDCG         |
|---------------|--------------|--------------|
| Content Based | 0.109        | 0.26         |
| Hybrid        | <b>0.161</b> | <b>0.323</b> |

Table 1: MRR@100 và NDCG@100

Kết quả của phương pháp MRR (Mean Reciprocal Rank) và NDCG (Normalized Discounted Cumulative Gain) thu được như bảng 1. Cả hai phương pháp, lọc nội dung và phương pháp lai, đều cho thấy hiệu suất tương đối ổn định. Phương pháp lọc nội dung có giá trị MRR@100 và NDCG@100 đạt được tương ứng là 0.109 và 0.26. Trong khi

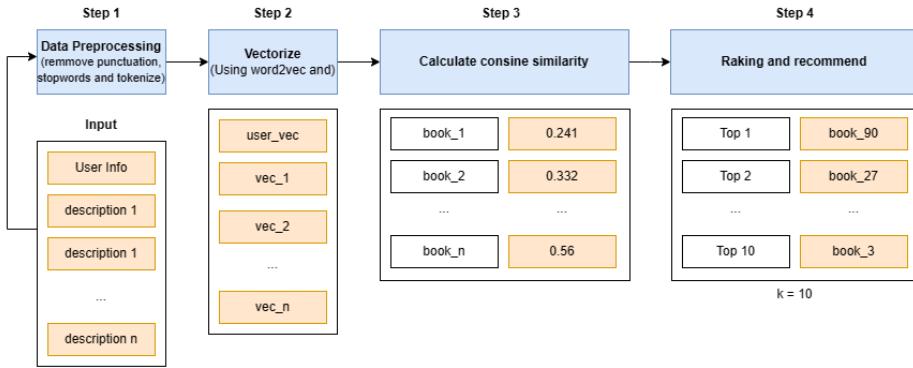


Figure 10: Quy trình khuyến nghị dựa trên phương pháp lọc nội dung

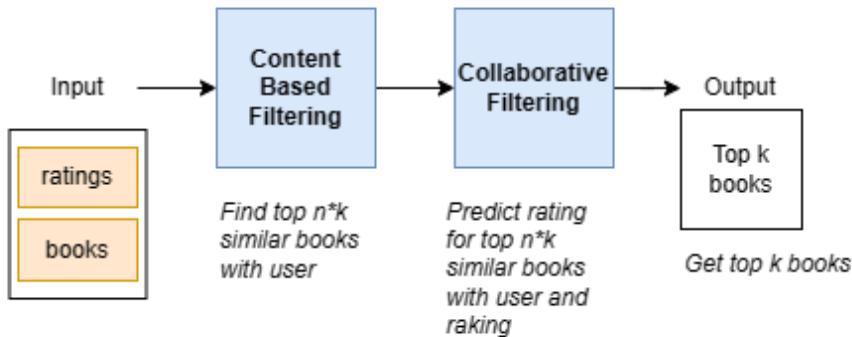


Figure 11: Quy trình khuyến nghị dựa trên phương pháp lai

đó, phương pháp lai có hiệu suất cao hơn với MRR@100 là 0.161 và NDCG@100 là 0.323. Kết quả này cho thấy sự độ vượt trội của phương pháp lai trên hầu hết các thang đo, làm nổi bật tính hiệu quả khi kết hợp các phương pháp. Tổng kết thu được từ bảng cho thấy cả hai phương pháp lọc nội dung và phương pháp lai đã đáp ứng tốt mục tiêu của nhóm về hiệu suất đề xuất. Sự độ vượt trội của phương pháp lai là một dấu hiệu tích cực, chứng minh rằng việc kết hợp các phương pháp có thể mang lại kết quả tốt hơn cho hệ thống khuyến nghị.

Đối với phương pháp lọc cộng tác, kết quả phân tích cho thấy cả hai phương pháp đều cho ra sai số khá cao. Tuy nhiên phương pháp lọc cộng tác dựa trên thuật toán ALS cho thấy kết quả tốt hơn so với thuật toán KNN (xem bảng 2).

| Phương pháp | RMSE        | MAE          | R2 score     |
|-------------|-------------|--------------|--------------|
| ALS         | <b>0.71</b> | <b>0.563</b> | <b>0.487</b> |
| KNN         | 0.914       | 0.72         | 0.15         |

Table 2: Phân tích kết quả ALS và KNN

Phương pháp cuối cùng là lọc dựa trên tri thức. Độ chính xác mà phương pháp này đưa ra là khá thấp sau khi phân tích trên test gồm 100 cuốn sách (xem bảng 3 ).

| Phương pháp     | Presision@100 | Recal@100 | F1@100 |
|-----------------|---------------|-----------|--------|
| Knowledge Based | 0.00112       | 0.07142   | 0.0102 |

Table 3: Phân tích kết quả Knowledge Based

## 6 Kết luận

Ở nghiên cứu này, chúng tôi đã triển khai thành công hệ thống khuyến nghị sách cung cấp các đề xuất cá nhân hóa cho người dùng. Hệ thống của chúng tôi tích hợp các phương pháp khác nhau như lọc cộng tác, lọc dựa trên tri thức, lọc dựa trên nhân khẩu học, lọc dựa trên nội dung và phương pháp lai. Điều này giúp tối ưu hóa hiệu suất của hệ thống, cung cấp đề xuất sách đa dạng và chính xác. Kết quả đạt được đã đáp ứng được mục tiêu của nhóm nghiên cứu. Hiệu suất của hệ thống tương đối ổn định. Trong tương lai, chúng tôi đặt kế hoạch cải

thiện độ chính xác của hệ thống bằng cách tối ưu hóa các phương pháp hiện tại và thử nghiệm với các phương pháp tiên tiến mới như mô hình mạng nơ-ron (NN models). Điều này giúp nâng cao trải nghiệm người dùng và làm cho hệ thống khuyến nghị trở nên ngày càng thông minh.

## Acknowledgements

Chúng tôi muốn cảm ơn những người đánh giá, những nhận xét của người hướng dẫn giúp chúng tôi cải thiện chất lượng bài nghiên cứu của mình.

Đầu tiên, nhóm nghiên cứu dành lời cảm ơn tới thầy Huỳnh Văn Tín đã hỗ trợ và đưa ra lời nhận xét cho nhóm trong suốt quá trình nghiên cứu. Chuyên môn của thầy đã giúp nhóm nghiên cứu định hướng và cải thiện thêm hệ khuyến nghị. Tiếp đến nhóm nghiên cứu muốn dành lời cảm ơn đến các website đã cung cấp tài nguyên cho nhóm nghiên cứu thực hiện nghiên cứu này. Cuối cùng, nhóm nghiên cứu dành lời cảm ơn đến bạn bè và gia đình đã hỗ trợ và tạo động lực cho nhóm nghiên cứu.

Mặc dù các cá nhân và tổ chức phía trên đóng vai trò quan trọng trong nghiên cứu của nhóm, tuy nhiên các sai sót trong nghiên cứu này điều hoàn toàn là lỗi của nhóm nghiên cứu.

## References

Denis Kotkov, Alan Medlar, Alexandr Maslov, Umesh Raj Satyal, Mats Neovius, and Dorota Glowacka. 2022. [The tag genome dataset for books](#). In *Proceedings of the 2022 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22)*.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. [Efficient estimation of word representations in vector space](#). In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, and Ion Stoica. 2010. Spark: Cluster computing with working sets. In *Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing (HotCloud)*.