

GỢI Ý MÓN ĂN KẾT HỢP KỸ THUẬT LỌC NỘI DUNG VÀ LỌC CỘNG TÁC THEO NGƯỜI DÙNG

DISH RECOMMENDATION COMBINING CONTENT-BASED AND USER-BASED COLLABORATIVE FILTERING TECHNIQUES

NGUYỄN THỊ BẢO HIỀN^{1,a}, NGUYỄN THÁI NGHE²,
TRƯƠNG MỸ THU THẢO¹

¹Trường Đại học Sư phạm Kỹ thuật Vĩnh Long;

²Khoa Hệ thống Thông tin, Đại học Cần Thơ.

^aTác giả liên hệ: nguyenthibaohien@gmail.com

Nhận bài(Received): 04/12/2024; Phản biện(Reviewed): 30/12/2024; Chấp nhận(Accepted): 03/02/2025

TÓM TẮT

Hiện nay, các hệ thống gợi ý đã thu hút được sự quan tâm lớn trong lĩnh vực nghiên cứu và ứng dụng thực tiễn. Việc áp dụng các kỹ thuật gợi ý vào lĩnh vực ẩm thực, đặc biệt là gợi ý món ăn, vẫn còn nhiều tiềm năng để khai thác và phát triển. Đề tài “Gợi ý món ăn kết hợp kỹ thuật lọc nội dung và lọc cộng tác theo người dùng” đề xuất mô hình cụ thể để giải quyết bài toán đặt ra dùng kỹ thuật lọc theo nội dung kết hợp với lọc cộng tác để gợi ý món ăn cho khách hàng. Tiếp theo là hệ thống được tích hợp vào menu điện tử để thực nghiệm, đưa ra những gợi ý một cách trực quan cho người dùng. Kết quả thực nghiệm của mô hình có giá trị sai số bình phương trung bình gốc (RMSE) là 0.325 và có giá trị F1-Score là 66.3%.

Từ khóa: Hệ thống gợi ý, gợi ý món ăn, lọc cộng tác, lọc nội dung.

ABSTRACT

Current recommender systems have attracted significant attention in research and practical applications. Applying recommendation techniques to the culinary field, especially in suggesting dishes, still holds considerable potential for exploration and development. This paper, titled “Dish Recommendation Combining Content-Based and User-Based Collaborative Filtering Techniques” proposes a specific model to address this problem. It employs a hybrid approach combining content-based filtering with collaborative filtering to recommend dishes to customers. Subsequently, the system is integrated into an electronic menu for practical experimentation, offering intuitive recommendations to users. The experimental results of the model show a root mean squared error (RMSE) of 0.325 and an F1-Score of 66.3%.

Keywords: Recommender system, dish recommendation, collaborative filtering, content-based filtering.

1. GIỚI THIỆU

1.1 Giới thiệu bài toán

Hệ thống gợi ý món ăn không chỉ đơn

giản là các ứng dụng phần mềm cung cấp

các đề xuất và đề xuất được cá nhân hóa

cho các món ăn dựa trên sở thích, nhu cầu

ăn kiêng của người dùng và các yếu tố liên quan khác. Với việc phát triển ngày càng tăng của các lựa chọn món ăn đa dạng, người dùng có thể bị choáng ngợp khi điều hướng qua các trang lựa chọn rộng rãi và đưa ra quyết định sáng suốt. Hệ thống gợi ý món ăn không chỉ là một trong những giải pháp tốt nhất trong môi trường này, mà còn bằng cách phân tích sở thích của người dùng, nhu cầu ăn kiêng và các yếu tố liên quan khác để đưa ra gợi ý phù hợp. Hệ thống đề xuất món ăn không chỉ giới thiệu món ăn cho người dùng dựa trên tên món ăn, id món ăn, loại món ăn, loại chế độ ăn kiêng như chay hoặc không chay trong trường hợp đề xuất lọc dựa trên nội dung. Để đề xuất món ăn với sự trợ giúp của tính năng lọc cộng tác, chúng tôi đã sử dụng id người dùng, id món ăn và xếp hạng làm thuộc tính.

1.2 Những nghiên cứu liên quan

Nghiên cứu của nhóm tác giả M. Trattner và D. Elswiler (2017) với đề tài “Food Recommendation: Framework, Existing Solutions, and Challenges” [6] đã tổng hợp các kỹ thuật phổ biến trong hệ thống gợi ý món ăn và nhấn mạnh rằng các phương pháp hiện tại, bao gồm lọc nội dung và lọc cộng tác, thường hoạt động kém khi xử lý các yếu tố như hạn chế dị ứng hoặc sở thích ăn uống cá nhân hóa. Kết quả cho thấy sự kết hợp giữa dữ liệu dinh dưỡng và sở thích người dùng có thể nâng cao hiệu quả của hệ thống gợi ý. Ngoài ra, tác giả Adadi, S. F. (2021) cũng có đề tài nghiên cứu “Personalized Food Recommendation Using Machine Learning: A Review” [7]. Nghiên cứu này cho thấy rằng việc sử dụng các kỹ thuật học sâu (deep learning) kết hợp với mô hình lọc cộng tác cải thiện đáng kể độ chính xác trong dự đoán. Một số mô hình hybrid được áp dụng cho tập dữ liệu lớn đã đạt được mức độ chính xác trên

85% khi gợi ý món ăn. Ngoài ra, nghiên cứu cũng nhấn mạnh vai trò của dữ liệu ngữ nghĩa (semantic data) trong việc cải thiện hệ thống gợi ý.

Nghiên cứu của tác giả Nguyễn Thái Nghe (2024) và các đồng sự thực hiện đề tài nghiên cứu một giải pháp ứng dụng trí tuệ nhân tạo trong gợi ý món ăn cho các nhà hàng[1]. Đề tài đề xuất ứng dụng trí tuệ nhân tạo (AI) trong hệ thống gợi ý món ăn, tích hợp kỹ thuật lọc cộng tác và luật kết hợp. Hệ thống tập trung cá nhân hóa gợi ý dựa trên lịch sử đặt món, sở thích và đặc điểm của khách hàng tại các nhà hàng. Kết quả độ chính xác của các gợi ý đạt 91% qua phản hồi của người dùng thực tế. Tác giả Trần Nguyễn Minh Thư và Huỳnh Quang Nghi (2016) thực hiện đề tài xây dựng hệ thống gợi ý thức ăn nhanh. Hệ thống tập trung vào gợi ý các món ăn nhanh dựa trên dữ liệu lịch sử mua hàng và đánh giá của khách hàng. Nghiên cứu sử dụng lọc cộng tác dựa trên mô hình láng giềng (Neighborhood-based Collaborative Filtering). Kết quả gợi ý món ăn nhanh chính xác hơn cho các đối tượng khách hàng thường xuyên. Tăng tỷ lệ tương tác của người dùng lên 20%. Độ hài lòng của người dùng đối với các gợi ý đạt mức 85%.

2. PHƯƠNG PHÁP ĐỀ XUẤT

2.1 Lọc nội dung (Content-Based Filtering)

2.1.1 Nguyên lý

Dựa trên các thông tin chi tiết về món ăn như tên, mô tả các thành phần (nguyên liệu), hệ thống sẽ gợi ý các món ăn tương tự dựa trên độ tương đồng giữa các đặc điểm này.

2.1.2 Cách hoạt động cụ thể

- Phân tích thông tin món ăn mà khách hàng đã yêu thích hoặc lựa chọn trước đó

(ví dụ: món “Cơm gà” có mô tả thành phần như “gạo, thịt gà, rau thơm”).

- Sử dụng các phương pháp tính toán độ tương đồng (như Cosine Similarity hoặc TF-IDF) để tìm các món ăn có mô tả và thành phần tương tự.

- Gợi ý các món ăn gần nhất với sở thích của người dùng.

- Ví dụ: Nếu khách hàng thích món “Phở bò” (gồm thịt bò, bánh phở, hành lá, nước dùng), hệ thống có thể gợi ý các món tương tự như “Bún bò Huế” hoặc “Hủ tiếu bò”.

2.2 Loại cộng tác (Collaborative Filtering)

2.2.1 Nguyên lý:

Hệ thống dựa trên sự tương tác giữa người dùng với món ăn để gợi ý. Nếu hai

người dùng có hành vi chọn món tương tự, họ có thể thích các món mà người còn lại đã chọn nhưng mình chưa thử.

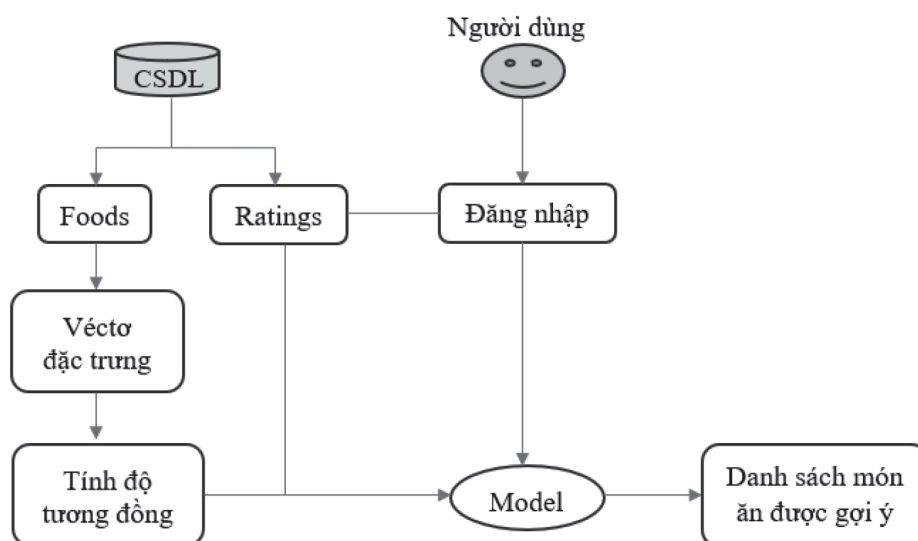
2.2.2 Cách hoạt động cụ thể:

- Dựa trên người dùng (User-based Collaborative Filtering): Tìm những người dùng có sở thích tương đồng với người dùng hiện tại, sau đó gợi ý món ăn dựa trên lịch sử chọn món của họ.

- Dựa trên sản phẩm (Item-based Collaborative Filtering): Phân tích mối quan hệ giữa các món ăn được chọn bởi các người dùng khác nhau để gợi ý các món phổ biến tương tự.

- Ví dụ: Nếu người dùng A thích “Pizza” và “Mi Ý”, còn người dùng B thích “Pizza” và “Hamburger”, hệ thống có thể gợi ý “Hamburger” cho A và “Mi Ý” cho B.

2.3 Mô hình tổng quan của hệ thống



Hình 1 Sơ đồ tổng quan hệ thống gợi ý

Trình tự các bước thực hiện:

- Bước 1: Chúng tôi đã thu thập thông tin liên quan như id người dùng, id món ăn và xếp hạng từ tập dữ liệu. Ngoài ra, chúng tôi đã thu thập tên món ăn liên quan đến từng id món ăn có độ tương

đồng về nội dung với các món ăn khách hàng đã rating.

- Bước 2: Sử dụng dữ liệu đã thu thập, chúng tôi đã tạo một ma trận bao gồm id người dùng, id món ăn và xếp hạng tương ứng. Ma trận này đóng vai trò là nền tảng

để nắm bắt các tương tác và sở thích của người dùng.

- Bước 3: Để xác định độ tương đồng giữa id người dùng và id món ăn, chúng tôi đã sử dụng độ tương đồng cosine. Ngoài việc xem xét xếp hạng, chúng tôi cũng kết hợp thông tin văn bản của tên món ăn bằng cách chuyển đổi chúng thành vector đặc trưng.

- Bước 4: Dùng phương pháp lọc theo nội dung, phân tích thông tin món ăn mà khách hàng đã lựa chọn trước đó (ví dụ: món “Cơm gà” có mô tả thành phần như “gạo, thịt gà, rau thơm”). Sử dụng các phương pháp tính toán độ tương đồng (Cosine Similarity) để tìm các món ăn có mô tả và thành phần tương tự. Gợi ý các món ăn gần nhất với sở thích của người dùng.

- Bước 5: Dùng phương pháp lọc cộng tác, dựa trên người dùng (User-based Collaborative Filtering): Tìm những người dùng có sở thích tương đồng với người dùng hiện tại, sau đó gợi ý món ăn dựa trên lịch sử chọn món của họ. Bằng cách phân tích mối quan hệ giữa các món ăn được chọn bởi các người dùng khác nhau để gợi ý các món phổ biến tương tự.

- Bước 6: Tiến hành kết hợp hai mô hình bằng phương pháp kết hợp muộn Weighted Average: Tính toán điểm gợi ý tổng hợp bằng cách cho trọng số khác nhau cho mỗi mô hình tùy thuộc vào độ tin cậy hoặc hiệu quả của mỗi mô hình.

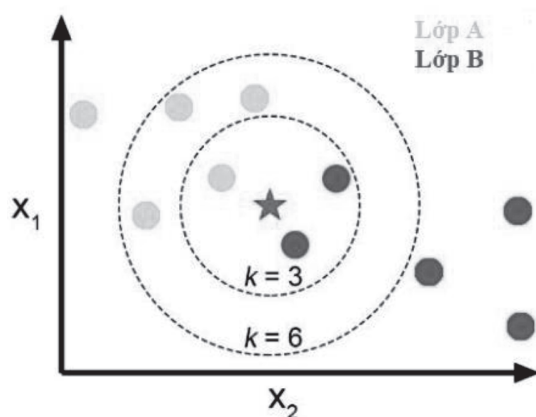
Bằng cách liên kết phương pháp lọc theo nội dung và phương pháp lọc cộng tác, chúng tôi đã nâng cao khả năng của hệ thống đề xuất để cung cấp các đề xuất món ăn được cá nhân hóa và nhận biết theo ngữ cảnh cho người dùng dựa trên nội dung các món ăn người dùng đã xếp hạng và thông tin văn bản liên quan đến các món ăn.

2.4 Tập dữ liệu

Đề tài gợi ý món ăn sử dụng tập dữ liệu do chúng tôi tự khảo sát và thu thập từ một số nhà hàng, cửa hàng ăn uống ở tỉnh Kiên Giang. Tập dữ liệu dùng để huấn luyện gồm 2 tệp tin csv (hình 3.1). Tệp tin Foods.csv gồm 5 trường: Food_ID, Name, C_Type, Veg_Non, Describe và 432 dòng là 432 món ăn. Tệp tin Ratings.csv gồm 3 trường: User_ID, Food_ID, Rating và 5000 dòng là thông tin khảo sát mức độ đánh giá của 500 lượt khách hàng về các món ăn của một khu ẩm thực tại Thành phố Hà Tiên, tỉnh Kiên Giang.

2.5 Kỹ thuật phân loại KNN

Phương pháp phân loại là một thuật toán tự động gán dữ liệu thành các danh mục hoặc lớp cụ thể. Nó sử dụng các quy tắc hoặc mẫu được xác định trước để thực hiện các phân loại này. Kết quả của quá trình học máy được áp dụng cho bộ phân loại là một mô hình phân loại. Mô hình này đóng vai trò là nền tảng để thực hiện phân loại dữ liệu. Trong hệ thống khuyến nghị được đề xuất, bộ phân loại K-láng giềng gần nhất được sử dụng [3.1]. Bằng cách sử dụng kỹ thuật này, chúng tôi cung cấp cho hệ thống dữ liệu và có được mô hình chính xác cho phép phân loại chính xác. Thuật toán láng giềng gần nhất là một cách tiếp cận đơn giản nhưng hiệu quả dựa trên các mẫu cục bộ trong dữ liệu để đưa ra dự đoán hoặc đề xuất. Đây là một thuật toán học không tham số, có nghĩa là nó không xây dựng mô hình một cách rõ ràng trong giai đoạn đào tạo mà trực tiếp sử dụng các trường hợp đào tạo để dự đoán. Trong thuật toán láng giềng gần nhất, độ tương đồng giữa các điểm dữ liệu được xác định bằng cách đo khoảng cách giữa chúng bằng thước đo khoảng cách đã chọn, chẳng hạn như khoảng cách Euclide hoặc độ tương đồng cosine như trong hình 2.



Hình 2 Minh họa thuật toán K-láng giềng

2.6 Độ tương đồng Cosine

Độ tương đồng cosine là thước đo được sử dụng để xác định độ tương tự giữa hai vectơ trong không gian nhiều chiều. Trong bối cảnh khuyến nghị về món ăn, độ tương tự cosine thường được áp dụng để tính toán độ tương tự giữa sở thích hoặc đặc điểm của các món ăn hoặc hồ sơ người dùng khác nhau. Độ tương tự cosine giữa hai vectơ A và B được tính bằng công thức sau:

$$\text{cosine similarity}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

A.B đại diện cho tích vô hướng của vectơ A và B, là tổng của phép nhân phần tử của các thành phần tương ứng của chúng.

$\|A\|$ và $\|B\|$ lần lượt biểu thị các chuẩn mực Euclide (còn được gọi là độ lớn hoặc độ dài) của vectơ A và B. Để áp dụng độ tương tự cosine trong khuyến nghị món ăn, vectơ A và B có thể biểu thị sở thích của người dùng hoặc đặc điểm của món ăn hoặc hồ sơ người dùng.

Các phần tử của vectơ có thể tương ứng với các thuộc tính như mùi vị, thành phần, giá trị dinh dưỡng hoặc bất kỳ đặc điểm liên quan nào khác. Bằng cách tính toán độ tương tự cosin giữa các món ăn hoặc hồ sơ người dùng, có thể xác định được sở thích

hoặc đặc điểm của họ giống nhau đến mức nào. Sau đó, thông tin tương tự này có thể được sử dụng để tạo ra các đề xuất, chẳng hạn như đề xuất các món ăn tương tự với những món ăn mà người dùng đã yêu thích trước đây hoặc xác định những người dùng có sở thích tương tự để đưa ra các đề xuất dựa trên lọc cộng tác.

2.7 Đánh giá độ tin cậy của giải thuật bằng thang đo RMSE

Root Mean Square Error (RMSE) hoặc Root Mean Square Deviation (RMSD) là căn bậc hai của mức trung bình của các sai số bình phương. RMSE là độ lệch chuẩn của các phần dư (sai số dự đoán).

Phần dư là thước đo khoảng cách từ các điểm dữ liệu đường hồi quy; RMSE là thước đo mức độ đàn trải của những phần dư này, nói cách khác, nó cho bạn biết mức độ tập trung của dữ liệu xung quanh đường phù hợp nhất.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}} \quad (1)$$

Trong đó:

RMSE là độ lệch bình phương gốc

i: Biến i

N: Số điểm dữ liệu không thiếu

x_i : Chuỗi thời gian quan sát thực tế

\hat{x}_i : Chuỗi thời gian ước tính

2.8 Phương pháp kết hợp có trọng số

Kết hợp hai phương pháp lọc nội dung (CBF) và lọc cộng tác (CF) được thực hiện thông qua một trọng số để điều chỉnh mức độ ảnh hưởng của mỗi phương pháp trong quá trình tạo ra gợi ý. Đề tài đã áp dụng trọng số kết hợp (Weighted Hybrid) với công thức như sau:

$$S_{\text{hybrid}}(u, i) = \alpha \cdot S_{CF}(u, i) + (1 - \alpha) \cdot S_{CBF}(u, i) \quad (2)$$

Trong đó:

- $S_{\text{hybrid}}(u, i)$: Điểm gợi ý cuối cùng cho người dùng u đối với sản phẩm i sau khi kết hợp cả CF và CBF.

- $S_{CF}(u, i)$: Điểm gợi ý từ phương pháp Content-Based Filtering đối với người dùng u và sản phẩm i .

- $S_{CBF}(u, i)$: Điểm gợi ý từ phương pháp Content-Based Filtering đối với người dùng u và sản phẩm i .

- α : Trọng số cho phương pháp Collaborative Filtering ($0 \leq \alpha \leq 1$).

- $(1-\alpha)$: Trọng số cho phương pháp

Content-Based Filtering.

3. KẾT QUẢ THỰC NGHIỆM

3.1 Thực nghiệm mô hình sử dụng phương pháp lọc theo nội dung

Kết quả thực nghiệm trên mô hình sử dụng phương pháp lọc theo nội dung dựa trên mô tả món ăn. Bằng cách chuyển đổi tên các món ăn thành các vectơ đặc trưng, so sánh độ tương đồng của món ăn mà khách hàng đã chọn với các món ăn còn lại để tìm món ăn gần giống gợi ý cho khách hàng. Kết quả thực nghiệm trên tập dữ liệu chia theo tỉ lệ 80% huấn luyện và 20% kiểm thử. Cho kết quả trên 10 lần chạy như sau:

Bảng 3.1 Kết quả 10 lần thực nghiệm trên mô hình lọc theo nội dung

| Lần thứ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Trung bình |
|---------------|------|------|------|------|------|------|------|------|------|------|------------|
| RMSE | 0.36 | 0.35 | 0.35 | 0.37 | 0.35 | 0.36 | 0.35 | 0.37 | 0.35 | 0.34 | 0.36 |
| Precision (%) | 61.2 | 61.3 | 61.4 | 61.2 | 61.2 | 61.2 | 61.5 | 61.2 | 61.4 | 61.2 | 61.3 |
| Recall (%) | 61.4 | 61.4 | 61.5 | 61.3 | 61.3 | 61.2 | 61.6 | 61.3 | 61.4 | 61.3 | 61.4 |
| F1-Score (%) | 61.3 | 61.3 | 61.4 | 61.2 | 61.2 | 61.2 | 61.5 | 61.2 | 61.4 | 61.2 | 61.3 |

Qua bảng kết quả sau 10 lần chạy thực nghiệm mô hình lọc nội dung cho ta thấy kết quả của các chỉ số tương đối cao.

RMSE: Các giá trị RMSE dao động từ 0.34 đến 0.37 trong 10 lần thực nghiệm, với giá trị trung bình là 0.36. Giá trị RMSE khá nhỏ và ổn định, cho thấy mô hình dự đoán có độ chính xác tương đối cao và ít biến động. Tuy nhiên, giá trị 0.36 cho thấy vẫn tồn tại một mức độ sai số nhất định giữa kết quả dự đoán của mô hình và thực tế.

Precision: Các giá trị Precision dao động từ 61.2% đến 61.5%, với giá trị trung bình là 61.3%. Độ chính xác của mô hình ở mức khá, cho thấy khoảng 61.3% số gợi

ý thức ăn mà mô hình dự đoán là phù hợp. Giá trị Precision không có nhiều sự biến động giữa các lần thực nghiệm, cho thấy tính ổn định của mô hình.

Recall: Các giá trị Recall dao động từ 61.2% đến 61.6%, với giá trị trung bình là 61.4%. Độ bao phủ của mô hình cũng ở mức khá, cho thấy mô hình có khả năng tìm ra khoảng 61.4% các loại thức ăn phù hợp trong tổng số các loại thức ăn phù hợp. Tương tự Precision, độ bao phủ cũng không có nhiều sự biến động, cho thấy tính ổn định của mô hình.

F1-Score: Các giá trị F1-Score dao động từ 61.2% đến 61.5%, với giá trị

trung bình là 61.3%. Điểm F1 trung bình là 61.3%, cho thấy hiệu quả tổng hợp của mô hình ở mức khá. Sự tương đồng về giá trị giữa Precision, Recall, và F1-Score cho thấy mô hình không có sự thiên lệch giữa khả năng tìm kiếm kết quả chính xác và khả năng bao phủ các kết quả phù hợp.

3.2 Thực nghiệm mô hình sử dụng phương pháp lọc cộng tác

Kết quả thực nghiệm trên mô hình sử

Bảng 3.2 Kết quả 10 lần thực nghiệm trên mô hình sử dụng phương pháp lọc cộng tác

| Lần thứ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Trung bình |
|---------------|------|------|------|------|------|------|------|------|------|------|------------|
| RMSE | 0.33 | 0.35 | 0.35 | 0.34 | 0.34 | 0.33 | 0.34 | 0.33 | 0.33 | 0.33 | 0.34 |
| Precision (%) | 63.3 | 62.5 | 62.7 | 63.6 | 64.8 | 63.6 | 62.8 | 63.7 | 64.4 | 62.9 | 63.4 |
| Recall (%) | 63.4 | 64.7 | 63.5 | 64.3 | 65.6 | 64.8 | 63.9 | 62.6 | 64.9 | 63.3 | 64.1 |
| F1-Score (%) | 63.3 | 63.6 | 63.1 | 63.9 | 65.2 | 64.2 | 63.3 | 63.1 | 64.6 | 63.1 | 63.8 |

Qua bảng kết quả sau 10 lần chạy thực nghiệm mô hình lọc cộng tác cho ta thấy kết quả có phần cải thiện hơn so với mô hình lọc theo nội dung.

RMSE: Các giá trị RMSE dao động từ 0.33 đến 0.37, với giá trị trung bình là 0.34. So với giá trị trung bình 0.36 ở kết quả trước, giá trị RMSE trung bình bây giờ đã giảm xuống 0.34. Điều này cho thấy, mô hình đã được cải thiện và dự đoán chính xác hơn. Sự dao động của RMSE vẫn khá ổn định trong khoảng 0.04, chứng tỏ tính nhất quán của mô hình.

Precision: Các giá trị Precision dao động từ 62.5% đến 64.8%, với giá trị trung bình là 63.4%. Độ chính xác trung bình đã tăng từ 61.3% (ở kết quả trước) lên 63.4%. Điều này cho thấy mô hình đưa ra gợi ý thức ăn chính xác hơn. Precision vẫn duy trì tính ổn định, không có sự biến động lớn giữa các lần thực nghiệm.

Recall: Các giá trị Recall dao động từ

dùng phương pháp lọc cộng tác, để tạo đề xuất cho từng id người dùng, chúng tôi đã tận dụng xếp hạng và tên món ăn của những người có sở thích món ăn gần giống nhất. Chúng tôi tổng hợp xếp hạng các món ăn của những người có sở thích gần giống nhất. Kết quả thực nghiệm trên tập dữ liệu chia theo tỉ lệ 80% huấn luyện và 20% kiểm thử. Cho kết quả trên 10 lần chạy như sau:

62.6% đến 65.6%, với giá trị trung bình là 64.1%. Độ bao phủ trung bình đã tăng từ 61.4% (ở kết quả trước) lên 64.1%. Điều này cho thấy mô hình có khả năng tìm ra nhiều hơn các loại thức ăn phù hợp. Recall cũng không có sự biến động lớn giữa các lần thực nghiệm, cho thấy tính ổn định của mô hình.

F1-Score: Các giá trị F1-Score dao động từ 63.1% đến 65.2%, với giá trị trung bình là 63.8%. Điểm F1 trung bình đã tăng từ 61.3% (ở kết quả trước) lên 63.8%, cho thấy hiệu quả tổng hợp của mô hình đã được cải thiện đáng kể. Sự tương đồng về giá trị giữa Precision, Recall, và F1-Score vẫn cho thấy mô hình không có sự thiên lệch giữa khả năng tìm kiếm kết quả chính xác và khả năng bao phủ các kết quả phù hợp.

3.3 Thực nghiệm mô hình kết hợp

Tiến hành kết hai mô hình với các trọng số khác nhau theo công thức (2) ở trên, kết quả thực nghiệm như sau:

Bảng 3.3 Kết quả nghiệm trên mô hình kết hợp với các trọng số khác nhau

| α | 0.5 | 0.45 | 0.4 | 0.35 | 0.3 | 0.25 | 0.2 |
|----------------------|-------|-------|-------|--------------|-------|-------|-------|
| RMSE | 0.353 | 0.347 | 0.336 | 0.321 | 0.326 | 0.330 | 0.357 |
| Precision (%) | 63.5 | 64.6 | 65.6 | 66.2 | 65.7 | 64.4 | 63.9 |
| Recall (%) | 64.7 | 64.9 | 65.3 | 66.5 | 65.6 | 64.9 | 64.3 |
| F1-Score (%) | 64.1 | 64.7 | 65.4 | 66.3 | 65.6 | 64.6 | 64.1 |

Qua thực nghiệm cho thấy kết quả kết hợp cao nhất với trọng số $\alpha=0.35$ là phù hợp nhất, giá trị RMSE = 0.32

RMSE: RMSE giảm dần khi α giảm từ 0.5 đến 0.35, sau đó có xu hướng tăng lên khi α tiếp tục giảm. Giá trị RMSE nhỏ nhất là 0.321, đạt được khi $\alpha = 0.35$. Điều này cho thấy, trọng số $\alpha = 0.35$ giúp mô hình dự đoán chính xác nhất, ít sai số nhất.

Precision: Độ chính xác tăng khi α giảm từ 0.5 đến 0.35, sau đó có xu hướng giảm xuống khi α tiếp tục giảm. Độ chính xác cao nhất là 66.2%, đạt được khi $\alpha = 0.35$. Điều này cho thấy khi $\alpha = 0.35$, mô hình có khả năng gợi ý thức ăn phù hợp nhất so với tổng số gợi ý đưa ra.

Recall: Độ bao phủ tăng khi α giảm từ 0.5 đến 0.35, sau đó giảm khi α tiếp tục giảm. Độ bao phủ cao nhất là 66.5%, đạt được khi $\alpha = 0.35$. Tương tự Precision, $\alpha = 0.35$ giúp mô hình có khả năng tìm ra nhiều hơn các loại thức ăn phù hợp trong tổng số các loại thức ăn phù hợp.

F1-Score: Điểm F1 tăng khi α giảm từ 0.5 đến 0.35, sau đó giảm khi α tiếp tục giảm. Điểm F1 cao nhất là 66.3%, đạt được khi $\alpha = 0.35$. Điểm F1 cao nhất khi $\alpha = 0.35$ chứng tỏ, trọng số này mang lại sự cân bằng tốt nhất giữa Precision và Recall, đồng nghĩa với hiệu quả tổng hợp cao nhất của mô hình.

Kết luận:

- Trọng số $\alpha = 0.35$ cho thấy hiệu quả tốt nhất cho mô hình kết hợp, với RMSE thấp nhất (0.321) và F1-Score cao nhất (66.3%).

- Trọng số 0.35 cũng cho thấy sự cân bằng tốt giữa Precision (66.2%) và Recall (66.5%), cho thấy mô hình hoạt động ổn định về cả hai mặt chính xác và bao phủ.

- Các chỉ số đều cao hơn so với kết quả trước đó, chứng tỏ việc kết hợp hai mô hình với trọng số phù hợp đã cải thiện đáng kể hiệu quả gợi ý thức ăn.

- Việc thay đổi trọng số α có ảnh hưởng đáng kể đến hiệu quả của mô hình. Việc tìm được giá trị α tối ưu là rất quan trọng để đạt hiệu quả tốt nhất.

3.4 Tiến hành thực nghiệm với mô hình đã huấn luyện

Tiến hành thực hiện chạy Demo hệ thống, bằng cách chọn ngẫu nhiên ID khách hàng trong danh sách các khách hàng trong bảng Rating để hệ thống gợi ý.

Hình 4 minh họa khách hàng có mã ID là 72; 5 dòng đầu tiên hiển thị 5 món ăn mà khách hàng này đánh giá số điểm cao nhất. 10 dòng cuối hiển thị 10 món ăn mà hệ thống gợi ý dựa trên những món ăn của các khách hàng có mức đánh giá các món ăn tương đồng với khách hàng có mã ID là 72.

```

Showing recommendations for users: 72
=====
Food with high ratings from user
-----
Bánh gai: Bánh (chay)
Canh tần ô chả cá: Canh (mặn)
Cá lóc kho khế: Kho (mặn)
Nước ép bưởi: Nước giải khát (chay)
Chân gà nướng mật ong: Nướng (mặn)
-----
Top 10 food recommendation
-----
Mực rim me: Ăn vặt (mặn)
Canh măng giò heo: Canh (mặn)
Canh rau ngót thịt băm: Canh (mặn)
Cá trạch chiên giòn: Chiên (mặn)
Bánh canh ghe: Điểm tâm (mặn)
Cá ngán kho cà: Kho (mặn)
Chè đậu đen: Tráng miệng (Ngọt) (chay)
Chè hoa cau: Tráng miệng (Ngọt) (chay)
Kem đậu đỏ: Tráng miệng (Ngọt) (chay)
Kem khoai môn: Tráng miệng (Ngọt) (chay)

```

Hình 3 Demo hệ thống với khách hàng có ID ngẫu nhiên

4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

4.1 Kết luận

Kết quả thử nghiệm trên mô hình gợi ý cho thấy hệ thống đạt được giá trị RMSE (Root Mean Square Error) là 0.325. Đây là một chỉ số khá ấn tượng, phản ánh rằng độ sai lệch giữa dự đoán của mô hình và đánh giá thực tế của người dùng đối với các món ăn là tương đối thấp. RMSE là một chỉ số quan trọng trong việc đánh giá độ chính xác của các mô hình gợi ý, và với giá trị này, có thể khẳng định rằng mô hình đã có thể dự đoán khá chính xác sở thích của người dùng.

Tuy nhiên, giá trị F1-Score của mô hình chỉ đạt 66.3%, một con số không được cao so với kỳ vọng. F1-Score là chỉ số đánh giá độ chính xác và khả năng thu hồi của hệ thống, đặc biệt là trong việc gợi ý các món ăn phù hợp với sở thích của người dùng. Dù kết quả này cho thấy hệ thống có thể gợi ý đúng các món ăn theo sở thích của người dùng ở mức độ nhất định, nhưng hiệu quả thực tế của hệ thống trong việc đáp ứng đúng nhu cầu của người dùng chưa thật sự cao. Điều này chỉ ra rằng mô hình vẫn còn

những hạn chế, đặc biệt trong việc tối ưu hóa độ chính xác của các gợi ý thực tế mà người dùng mong đợi.

Mặc dù hệ thống đạt được một mức độ chính xác khá tốt về mặt dự đoán đánh giá (thể hiện qua RMSE), nhưng hiệu quả tổng thể của mô hình, đặc biệt là về độ chính xác và khả năng thu hồi (thể hiện qua F1-Score), vẫn còn hạn chế. Sự chênh lệch giữa hai chỉ số này chỉ ra rằng mô hình vẫn còn khó khăn trong việc đưa ra những gợi ý hoàn toàn phù hợp với sở thích của người dùng trong một phạm vi rộng lớn các lựa chọn món ăn.

4.2 Hướng phát triển

Tối ưu hóa dữ liệu đầu vào và cải thiện chất lượng dữ liệu: Tập trung vào thu thập dữ liệu đầy đủ và chất lượng cao hơn về thói quen ăn uống, sở thích cá nhân, và đánh giá của người dùng để mô hình học tốt hơn và đưa ra gợi ý chính xác hơn.

Tích hợp thêm dữ liệu ngữ cảnh: Bổ sung thông tin ngữ cảnh như thời gian, địa điểm, thời tiết, và sức khỏe cá nhân để làm phong phú thêm dữ liệu đầu vào, từ đó cung cấp gợi ý phù hợp hơn.

Nâng cao khả năng cá nhân hóa bằng cách học hỏi từ phản hồi của người dùng: Phát triển hệ thống tự động cập nhật và điều chỉnh mô hình dựa trên phản hồi thực tế từ người dùng sau mỗi lần sử dụng, giúp cải thiện độ chính xác của gợi ý theo thời gian. Hoặc sử dụng phương pháp học tăng cường (Reinforcement Learning) bằng cách áp dụng kỹ thuật học tăng cường để mô hình có thể tự điều chỉnh dựa trên phản hồi của người dùng và tối ưu hóa các quyết định gợi ý.

TÀI LIỆU THAM KHẢO

- [1] Đ. T. Liên, “Hợp nhất lọc cộng tác và lọc nội dung bằng phương pháp học bán giám sát,” Trường Đại học Bưu chính Viễn thông, 2020.
- [2] Nguyễn Hùng Dũng, Nguyễn Thái Nghe. 2014. Hệ thống gợi ý sản phẩm trong bán hàng trực tuyến sử dụng kỹ thuật lọc cộng tác. Tạp chí Khoa học Trường Đại học Cần Thơ, số 31a (2014), trang 36-51. ISSN: 1859-2333.
- [3] N. T. Nghe, Đ. H. H. Nguyễn, T. Q. Toanh, và N. H. Hòa, “Một giải pháp ứng dụng trí tuệ nhân tạo trong gợi ý món ăn cho các nhà hàng,” Tạp chí Khoa học Đại học Cần Thơ, vol. 60, no. SDMD, pp. 17-24, 2024.
- [4] P. M. Chuân, L. T. Hương, T. Đ. Khang, and N. V. Hậu, “Hệ thống gợi ý sử dụng thuật toán tối ưu bầy đàn”. Kỷ yếu Hội nghị quốc gia lần thứ VIII “Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin” (FAIR 2015)
- [5] Nguyễn Tấn Phong, Nguyễn Thái Nghe. 2014. Một giải pháp trong xây dựng Hệ thống gợi ý bài hát. Trang 149-154, kỷ yếu hội thảo quốc gia lần thứ XVII: Một số vấn đề chọn lọc của CNTT&TT (@2014). Nhà xuất bản Khoa học và kỹ thuật. ISBN: 978-604-67-0426-3
- [6] R. Baeza-Yates and B. Ribeiro-Neto (1999), “Modern Information Retrieval”. Addison-Wesley.
- [7] T. Mitchell (1997), “Machine Learning”, 1 ed. McGraw Hill.
- [8] M. Pazzani and D. Billsus (1997), “Learning and Revising User Profiles: The Identification of Interesting Web Sites”, Machine Learning, vol. 27, pp. 313-331.
- [9] G. Somlo and A. Howe (2001), “Adaptive Lightweight Text Filtering”, Proc. Fourth Int’l Symp. Intelligent Data Analysis.
- [10] Y. Zhang and J. Callan (2001), “Maximum Likelihood Estimation for Filtering Thresholds”, Proc. 24th Ann. Int’l ACM SIGIR Conf.
- [11] G. Adomavicius, A. Tuzhilin (2005), “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions”, IEEE Transactions On Knowledge And Data Engineering, vol. 17, No. 6, 2005.
- [12] J.L. Herlocker, J.A. Konstan, and J. Riedl (2000), “Explaining Collaborative Filtering Recommendations”, Proc. ACM Conf. Computer Supported Cooperative Work.
- [13] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl (2001), “Item-Based Collaborative Filtering Recommendation Algorithms”, Proc. 10th Int’l WWW Conf.
- [14] J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J.T. Riedl (2004), “Evaluating Collaborative Filtering Recommender Systems”, ACM Trans. Information Systems, vol. 22, No. 1, pp. 5-53.
- [15] B. Marlin (2003), “Modeling User Rating Profiles for Collaborative Filtering”, Proc. 17th Ann. Conf. Neural Information Processing Systems (NIPS ’03).
- [16] D. Billsus and M. Pazzani (1998), “Learning Collaborative Information Filters”, Proc. Int’l Conf. Machine Learning.
- [17] D. Billsus and M. Pazzani (2000), “User Modeling for Adaptive News Access”, User Modeling and User-Adapted Interaction, vol. 10, No. 2-3, pp. 147-180

- [18] [103] X. Su and T. M. Khoshgoftaar (2006), “Collaborative filtering for multi-class data using belief nets algorithms”, in Proceedings of the International Conference on Tools with Artificial Intelligence (ICTAI '06), pp. 497–504.
- [19] X. Su, T. M. Khoshgoftaar (2009), “A Survey of Collaborative Filtering Techniques”. Advances in Artificial Intelligence, vol 2009, pp.1-20.
- [20] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl (2000), “Application of Dimensionality Reduction in Recommender Systems—A Case Study”, Proc. ACM WebKDD Workshop.
- [21] L.H. Ungar and D.P. Foster (1998), “Clustering Methods for Collaborative Filtering”, Proc. Recommender Systems, Papers from 1998 Workshop, Technical Report WS-98-08 1998.
- [22] L. Si and R. Jin (2003), “Flexible Mixture Model for Collaborative Filtering”, Proc. 20th Int'l Conf. Machine Learning.
- [23] T. Hofmann (2003), “Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis”, Proc. 26th Ann. Int'l ACM SIGIR Conf.
- [24] T. Hofmann (2004), “Latent Semantic Models for Collaborative Filtering”, ACM Trans. Information Systems, vol. 22, No. 1, pp. 89-115.
- [25] C. Desrosiers, G. Karypis (2008), “Solving the Sparsity Problem: Collaborative Filtering via Indirect Similarities”, Department of Computer Science and Engineering University of Minnesota (Technical Report).
- [26] D. Pavlov and D. Pennock (2002), “A Maximum Entropy Approach to Collaborative Filtering in Dynamic, Sparse, High-Dimensional Domains”, Proc. 16th Ann. Conf. Neural Information Processing Systems (NIPS '02).
- [27] C.C. Aggarwal, J.L. Wolf, K.L. Wu, and P.S. Yu (1999), “Hortling Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering”, Proc. Fifth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining.
- [28] Z. Huang, H. Chen, D. Zeng (2004), “Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering”, ACM Transactions on Information Systems, vol. 22(1) pp. 116–142
- [29] I. Soboroff and C. Nicholas (1999), “Combining Content and Collaboration in Text Filtering” Proc. Int'l Joint Conf. Artificial Intelligence Workshop: Machine Learning for Information Filtering.
- [30] C. Basu, H. Hirsh, and W. Cohen (1998), “Recommendation as Classification: Using Social and Content-Based Information in Recommendation”, Recommender Systems. Papers from 1998 Workshop, Technical Report WS-98-08, AAAI Press 1998.
- [31] J. Balisico, T. Hofmann (2004), “Unifying collaborative and content-based filtering”, In Proceedings. of Int'l. Conf. on Machine learning (ICML-2004).
- [32] A.I. Schein, A. Popescul, L.H. Ungar, and D.M. Pennock (2002), “Methods and Metrics for Cold-Start Recommendations”. Proc. 25th Ann. Int'l ACM SIGIR Conf.
- [33] P. Melville, R.J. Mooney, and R. Nagarajan (2002), “Content-Boosted Collaborative Filtering for Improved Recommendations”, Proc. 18th Nat'l Conf. Artificial Intelligence.
- [34] A. Gunawardana, C. Meek (2009), “A unified approach to building hybrid recommender systems. Microsoft Research , RecSys 2009: 117-124.