FOOD RECOMMENDATION SYSTEM

Hồ Thanh Duy Khánh, Nguyễn Hiếu Nghĩa, Nguyễn Thị Nguyệt, Ngô Thị Phúc, Huỳnh Văn Tín
Faculty of Information Science and Engineering, University of Information Technology,
Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
{20521445,20521654,20521689,20521765}gm.uit.edu.vn
tinhv@uit.edu.vn

Abstract

Trong bối cảnh cuộc cách mạng công nghiệp 4.0 đang chuyển đổi mọi khía cạnh của cuộc sống, việc áp dụng và nâng cao các hệ thống đề xuất (Recommendation System) đã đạt được thành công trong việc gợi ý các sản phẩm như phim, âm nhạc, hình ảnh, quần áo,..Hệ thống đề xuất món ăn không chỉ đơn giản là một công cụ hữu ích, mà còn là người bạn đồng hành đặc biệt trong thế giới ẩm thực ngày nay. Tuy nhiên, việc triển khai các hệ thống gợi ý trong đề xuất món nhận ít sự chú ý hơn. Vì vậy, đồ án mà nhóm chúng tôi sẽ thực hiện áp dụng và nâng cao các phương pháp hệ thống đề xuất món ăn. Với sự kết hợp linh hoạt giữa các phương pháp kỹ thuật trong hệ thống gợi ý thực phẩm, chúng tôi đề xuất các kỹ thuật trong bài báo này: phương pháp Collaborative filtering dựa trên thuật toán K-nearest neighbor (KNN), SVD, Neural Collaborative Filtering. Bên cạnh đó chúng tôi cũng áp dụng mô hình Ridge Regression trong phương pháp Content-based Filtering và cuối cùng phương pháp Hybrid Filtering sử dụng mô hình LightFM và SentimentNetwork, với mục đích đưa ra gợi ý dựa trên sở thích của người dùng, nhằm tránh quá tải thông tin dựa trên bộ dữ liệu mà chúng tôi xây dựng. Sự cải tiến các kỹ thuật nhằm mục đích cải thiện độ chính xác trong việc lựa chọn của người dùng và tạo ra những đề xuất phù hợp. Kết quả thực nghiệm cho thấy Nẻual Collaborative Filtering là phương pháp gợi ý tối ưu nhất và đạt được hiệu suất tốt nhất dựa trên các chỉ số đánh giá chất lượng của hệ thống gợi ý, vượt trội hơn so với các phương pháp dựa trên bộ nhớ truyền thống

1 Giới thiệu

Ẩm thực ngày nay đã trở thành một phần không thể thiếu trong cuộc sống của mọi người. Đối với nhiều người, ẩm thực không chỉ là vấn đề của nhu cầu ăn uống hàng ngày mà còn liên quan chặt chẽ đến lối sống cộng đồng. Ngày nay, việc ăn uống không chỉ đơn giản là thỏa mãn nhu cầu ẩm thực mà còn trở thành một phần của lối sống cộng đồng.

Với sự gia tăng sở thích ăn uống dẫn đến việc tìm kiếm thông tin và dữ liệu về các món ăn trở nên quá tải khiến người dùng không thể chọn chọn lọc chính xác theo sở thích của mình. Mặc dù thông tin từ các trang web hay ứng dụng vẫn sử dụng các bộ lọc dựa trên loại hình ẩm thực, điều này vẫn có khả năng hiển thị những lựa chọn phù hợp với kết quả lọc từ người dùng nhưng không chính xác với khẩu vị của họ.

Hầu hết người tiêu dùng thường chọn những món ăn có đánh giá cao và nhận xét tích cực từ người khác để đưa ra quyết định. Do đó, đánh giá và nhận xét đóng vai trò quan trọng đối với những người tiêu dùng có khẩu vị tương đồng, giúp họ nhận được các đề xuất tương tự từ những người dùng có khẩu vị giống nhau.

Trong nghiên cứu này, chúng tôi đặt ra một hệ thống đề xuất độc đáo, kết hợp linh hoạt giữa lọc dựa trên nội dung và kỹ thuật lọc cộng tác. Chúng tôi đã tiến hành phát triển và cải thiện hệ thống gợi ý món ăn bằng cách tích hợp các phương pháp phổ biến như Collaborative Filtering, Content-based Filtering, và Hybrid LightFM dựa trên người dùng. Phương pháp Collaborative Filtering, được sử dụng rộng rãi trong thiết kế hệ thống, tập trung vào việc đề xuất món ăn dựa trên sự tương đồng giữa sở thích của người dùng và nhóm người dùng khác. Bằng cách này, người dùng có thể nhận được các gợi ý từ những người có sở thích tương tự, tạo nên trải nghiệm gần gũi và chân thực. Phương pháp Content-based Filtering, ngược lại, tập trung vào đánh giá và gợi ý món ăn dựa trên các đặc trưng và thuộc tính của sản phẩm. Chúng tôi sử dụng thông tin chi tiết về món ăn, từ thành phần công thức đến phản hồi đánh giá từ người dùng, để tối ưu hóa đề xuất và đáp ứng đa dạng nhu cầu ẩm thực của người dùng. Hệ thống Hybrid LightFM là bước tiến đột phá, kết hợp sức mạnh của cả Collaborative và Content-based Filtering. Bằng cách này, chúng tôi tận dụng đánh giá từ người dùng để đào tạo và tối ưu hóa mô hình, tạo ra những đề xuất chính xác

và đa dạng. Quá trình lặp lại và điều chỉnh được thực hiện đều đặn để đảm bảo rằng hệ thống liên tục cải thiện và đáp ứng sở thích người dùng một cách tối ưu. Với sự kết hợp độc đáo này, chúng tôi hy vọng hệ thống của mình không chỉ mang lại trải nghiệm gợi ý món ăn chính xác mà còn đáp ứng đa dạng nhu cầu và mong muốn ẩm thực của người dùng, đặt chúng vào trung tâm của thế giới số hiện đại.

Nhiệm vụ chính của bài toán hệ khuyến nghị món ăn dựa trên đánh giá và phản hồi của người dùng và cho ra top N danh sách món ăn được đánh giá cao nhất tương tự với yêu cầu của người dùng, bài toán được miêu tả như sau:

- Đầu vào: Tên món ăn.
- Đầu ra: Top N danh sách món ăn đề xuất.

Phần còn lại của nghiên cứu được tổ chức như sau: chương hai chúng tôi xem xét các công trình đã có trong các hệ thống Gợi ý, Lọc Cộng tác, Gợi ý Dựa trên Nội dung, Gợi ý Kết hợp LightFM gồm sự kết hợp Lọc Cộng tác và Lọc Dựa trên Nội dung.Trong phần 3, chúng tôi giải thích quá trình xây dựng dữ liệu. Các phương pháp thực nghiệm được mô tả ở phần 4. Phần 5, chúng tôi trình bày các kết quả thực nghiệm và phân tích kết quả. Cuối cùng, phần 6 rút ra kết luận và công việc tương lai.

2 Các nghiên cứu liên quan

Công trình "A Hybridized Recommendation System on Movie Data Using Content-Based and Collaborative Filtering" là một nghiên cứu tại African University of Science and Technology, Abuja. Nghiên cứu này tập trung vào việc kết hợp lọc dựa trên nội dung và lọc cộng tác để cung cấp gợi ý phim cá nhân và đa dạng.

Nghiên cứu "Restaurant Recommender System Using User-Based Collaborative Filtering Approach: A Case Study at Bandung Raya Region" tập trung vào lĩnh vực đặc biệt của ứng dụng hệ thống gợi ý cho nhà hàng. Nó sử dụng phương pháp lọc cộng tác dựa trên người dùng để tối ưu hóa trải nghiệm ẩm thực.

Trong lĩnh vực thực phẩm, Hệ thống Tư vấn (RSs) đóng một vai trò quan trọng trong việc thúc đẩy các hành vi ăn uống lành mạnh bằng cách đề xuất các thay thế thực phẩm lành mạnh cho người dùng. Có thể chia Hệ thống Tư vấn thực phẩm thành ba loại dựa trên thông tin được sử dụng cho việc đề xuất thực phẩm.

Loại đầu tiên áp dụng sở thích ẩm thực của người dùng cho đề xuất, ví dụ, các thuật ngữ tìm kiếm hoặc thông tin nguyên liệu của người dùng đã được sử dụng để thực hiện đề xuất công thức nấu ăn. Loại thứ ba tìm kiếm sự cân bằng giữa sở thích của người dùng và nhu cầu dinh dưỡng, ví dụ, một kế hoạch ăn lành mạnh và dinh dưỡng đã được tạo ra cho người già trong bằng cách tận dụng thông tin từ cả sở thích của người dùng và dinh dưỡng thực phẩm.

Để cải thiện độ chính xác của Hệ thống Tư vấn thực phẩm, quan trọng là phải xem xét cả sở thích động của người dùng và phản hồi từ hàng xóm trong quá khứ. Trong chương này, Hệ thống Tư vấn dựa trên chuỗi (SRSs) được giới thiệu để nắm bắt sở thích động của người dùng. Mô hình hóa mẫu chuỗi hành vi của người dùng cho phép Hệ thống Tư vấn hiểu rõ sự phát triển của khẩu vị người dùng theo thời gian, từ đó cung cấp đề xuất tốt hơn [G.-E. Yap, X.-L. Li, and P. S. Yu, Effective next-items recommendation via personalized sequential pattern mining, In: Proceedings of the International Conference on Database Systems for Advanced Applications, Busan, South Korea, Apr. 2012, pp. 48-64].

3 Dữ liệu

3.1 Thu thập bộ dữ liệu

Chúng tôi phát triển một quy trình tự động hóa thu thập dữ liệu từ trang web Food.com, tập trung lấy thông tin về các công thức nấu ăn và đánh giá từ người dùng. Để thực hiện điều này, chúng tôi sử dụng hai công cụ chính là Selenium và BeautifulSoup, giúp tự động điều khiển trình duyệt và phân tích cú pháp HTML của trang web.

Quy trình bắt đầu bằng việc sử dụng Selenium để duyệt qua hàng loạt các trang trên Food.com. Sau đó, thu thập danh sách các công thức nấu ăn và URL tương ứng từ mỗi trang, nhằm có danh sách đầy đủ để lấy thông tin chi tiết về các món ăn. Tiếp theo, sử dụng Selenium để tự động truy cập từng URL của các công thức nấu ăn trên Food.com. Tại mỗi URL này, áp dụng các kỹ thuật của thư viện recipe-scrapers để thu thập thông tin chi tiết về tên món ăn, hình ảnh, thành phần, dinh dưỡng và đánh giá từ người dùng. Quá trình này được thực hiện tự động và liên tục từng trang để đảm bảo việc thu thập dữ liệu một cách toàn diện.

Sau khi thu thập thông tin từ mỗi trang, các dữ liệu này sau đó được cấu trúc lại và lưu trữ vào các cấu trúc dữ liệu. Kết quả cuối cùng là việc lưu trữ dữ liệu vào các tệp CSV để tiện cho việc sử dụng và phân tích dữ liệu.

3.2 Tiền xử lý dữ liệu

3.2.1 Tiền xử lý cơ bản

Tập dữ liệu thu thập từ quá trình crawl data trên trang Food.com sẽ trải qua quá trình tiền xử lý để chuẩn hóa và chuẩn bị dữ liệu cho các công đoạn tiếp theo của quá trình phân tích. Các bước tiền xử lý dữ liệu bao gồm:

- **Xử lý dữ liệu cơ bản:** Dữ liệu ban đầu về thành phần và dinh dưỡng được chuyển đổi từ dạng chuỗi sang dữ liệu có cấu trúc hợp lý; Loại bỏ các bản ghi trùng lặp dựa trên cột chứa đường dẫn của mỗi món ăn để đảm bảo tính duy nhất và chính xác của dữ liệu.

- **Chuẩn hóa dữ liệu Thông tin món ăn và dữ liệu Đánh giá:** Xử lý dữ liệu 'tên', 'thành phần' và 'dinh dưỡng' tạo thành một bảng thông tin về món ăn hoàn chỉnh; Dữ liệu về dinh dưỡng được chuyển đổi sang dạng số và đổi tên các cột để quản lý dễ dàng hơn. Dữ liệu rating và đánh giá văn bản từ người dùng cũng được xử lý để tạo thành một bảng thông tin về đánh giá món ăn.

- **Lưu trữ dữ liệu:** Kết quả sau khi xử lý được lưu xuống 2 file CSV khác nhau, bao gồm Thông tin về món ăn ("food.csv": Id_food, Name, Ingredients, Calories, Fat, Saturated_fat, Cholesterol, Sodium, Carbohydrate, Fiber, Sugar, Protein) và Đánh giá từ người dùng ("rating.csv": Id_user, Id_food, Rating, Review).

3.2.2 Tiền xử lý

Với các bước chuẩn hóa và xử lý cơ bản trên, quy trình tiền xử lý cuối cùng cho dữ liệu bao gồm các bước chính như sau:

- **Gộp dữ liệu:** Dữ liệu từ hai bảng thông tin về món ăn và đánh giá từ người dùng đã được gộp lại để tạo thành một bảng duy nhất.

- **Xử lý dữ liệu bị khuyết:** Dữ liệu thiếu (nếu có) được loại bỏ để đảm bảo tính chính xác và đồng nhất của dữ liệu trong quá trình phân tích.

- **Ánh xạ ID:** Các ID của người dùng và món ăn được chuyển đổi thành số nguyên liên tục từ 1 để đơn giản hóa việc xử lý. Ví dụ, ánh xạ các ID từ 'f_123' thành số nguyên 123.

- **Tiền xử lý văn bản:** Các cột chứa thông tin về nguyên liệu và đánh giá từ người dùng được tiền xử lý để chuẩn hóa dữ liệu. Các phương pháp tiền xử lý được áp dụng: Chuẩn hóa chữ hoa thành chữ thường, mở rộng từ viết tắt, loại bỏ các ký tự đặc biệt, số, dấu câu và từ ngữ không cần thiết (stopword), ngoài ra trong 'ingredients' còn loại bỏ các đơn vị đo lường như 'g', 'ml', 'teaspoon',... để tập trung vào thông tin vào các nguyên liệu chính.

- **Lọc dữ liệu:** Chỉ giữ lại những mẫu dữ liệu có đủ lượng thông tin để phân tích, bao gồm: Các mẫu dữ liệu mà món ăn đó có nhiều hơn 50 lượt rating từ người dùng, và các mẫu dữ liệu mà người dùng đã rating/review cho ít nhất 20 món ăn khác nhau.

Quy trình tiền xử lý này giúp chuẩn bị dữ liệu một cách toàn diện và chính xác, tạo điều kiện thuận lợi cho việc tiếp tục xây dựng mô hình và phân tích dữ liệu sau này.

3.3 Phân tích bộ dữ liệu

3.3.1 Dữ liệu sau bước tiền xử lý cơ bản

Sau quá trình xử lý cơ bản, dữ liệu hiện có 143545 người dùng, 9547 công thức/món ăn, và 479452 lượt rating và review.

No.	Thuộc tính	Ý nghĩa
1	id_food	ID duy nhất cho mỗi món ăn
2	name_food	Tên của món ăn
3	ingredients	Thành phần chính của món ăn
4	calories	Lượng calo trong món ăn
5	fat	Lượng chất béo trong món ăn
6	saturated_fat	Lượng chất béo no trong món ăn
7	cholesterol	Lượng cholesterol trong món ăn
8	sodium	Lượng natri trong món ăn
9	carbohydrate	Lượng carbohydrate trong món ăn
10	fiber	Lượng chất xơ trong món ăn
11	sugar	Lượng đường trong món ăn
12	protein	Lượng protein trong món ăn

Table 1: Thuộc tính của bộ dữ liệu food.csv

No.	Thuộc tính	Ý nghĩa
1	id_food	ID duy nhất cho mỗi món ăn
2	id_user	ID của người dùng
3	rating	Điểm đánh giá
4	reviews	Nhận xét văn bản từ người dùng về món ăn

Table 2: Thuộc tính của bộ dữ liệu rating.csv

3.3.2 Bộ dữ liệu

Bộ dữ liệu này chứa thông tin về món ăn và đánh giá món ăn từ 2323 người dùng, với 2838 công thức/món ăn, 110580 dòng, và 7 cột:

No.	Thuộc tính	Ý nghĩa
1	id_food	ID của món ăn, số nguyên từ 1 đến 2838
2	id_user	ID của người dùng, số nguyên từ 1 đến 2323
3	name_food	Tên của món ăn
4	rating	Điểm đánh giá
5	review_user	Nhận xét của người dùng về món ăn
6	ingredients	Thành phần nguyên liệu của món ăn
7	nutrients	Thông tin dinh dưỡng của món ăn

Table 3: Mô tả các thuộc tính của bộ dữ liệu

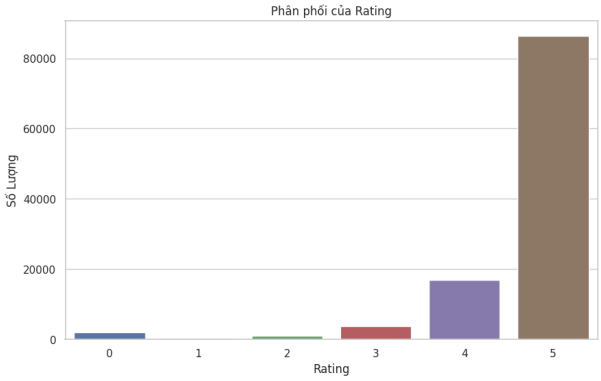

Figure 1: Số lượng rating của người dùng.

Dựa trên thống kê mô tả cơ bản, thấy rằng, đánh giá (rating) là giá trị số nguyên trong khoảng từ 0 đến 5, có độ lệch chuẩn là 0.88, có phân phối không đồng đều với số lượng đánh giá 5 (cao nhất) chiếm phần lớn, đến 78% tổng số lượng đánh giá, điều này cho thấy xu hướng tích cực trong các đánh giá.

Dựa vào biểu đồ hình 2, chúng ta có thể thấy xu hướng tăng lên từ biểu đồ phân tán, cho thấy những công thức nấu ăn nổi tiếng nhận được xếp hạng cao hơn. Phân phối xếp hạng trung bình cho thấy hầu hết các công thức nấu ăn trong tập dữ liệu đều có xếp hạng trung bình khoảng 4.6. Số lượng phân bổ rating cho thấy hầu hết mỗi công thức nấu ăn đều có lượng rating dưới 100.

Biểu đồ hình 3 cho thấy phân bố của điểm cảm xúc trên các đánh giá văn bản của người dùng, sử dụng TextBlob để tính điểm cảm xúc. Điểm cảm xúc được biểu diễn trên một thang từ -1 đến +1, với -1 là cực kỳ tiêu cực, 0 là trung lập, và +1 là cực kỳ tích cực. Từ các đánh giá văn bản, ta có thể thấy rằng phần lớn đánh giá có xu hướng tích cực nhẹ, với điểm cảm xúc tập trung chủ yếu trong khoảng từ 0 đến 0.5. Một số ít đánh giá có điểm

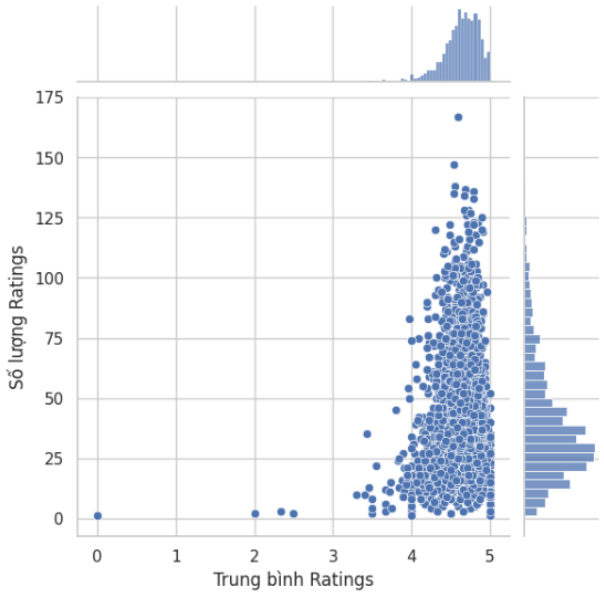

Figure 2: Mối tương quan giữa xếp hạng trung bình và số lượng xếp hạng

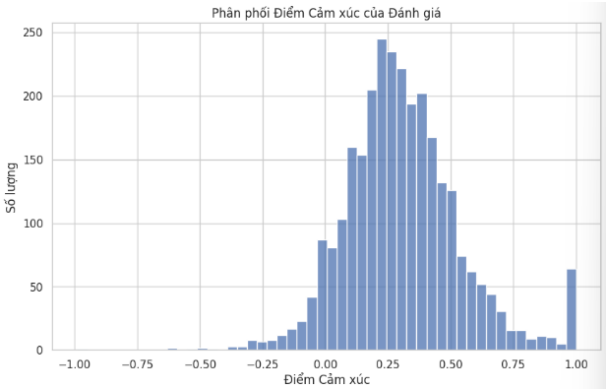

Figure 3: Phân phối điểm cảm xúc của review

cảm xúc tiêu cực, nhưng không nhiều. Điều này có thể cho thấy rằng đa số người dùng cảm thấy hài lòng đến mức độ nào đó với món ăn được đánh giá. Tuy nhiên, cũng có một lượng đáng kể các đánh giá trung lập, cho thấy không ít người dùng không có xu hướng cảm xúc mạnh mẽ về mặt tích cực hay tiêu cực. Điều này đồng nghĩa các đánh giá văn bản (điểm cảm xúc: tập trung nhiều trong khoảng 0-0.5) và đánh giá dựa trên số (rating: 5 chiếm 78%) chưa thực sự tương thích với nhau.

Dữ liệu này sau đó được chia thành 2 bộ dữ liệu là tập train và test theo tỷ lệ 8:2, đảm bảo rằng mỗi user có mặt trong tập train cũng có trong tập test (số lượng user trong tập train và tập test là 2323). Train, test có số chiều lần lượt là (88494, 7) và (22086, 7).

4 Phương pháp và cài đặt thực nghiệm

Trong nghiên cứu này, nhóm thực hiện 3 cách tiếp cận trong việc xây dựng hệ thống khuyến nghị thực phẩm:

(1) Dựa trên lọc cộng tác (collabrative filtering)
(2) Dựa trên nội dung (content-based filtering)
(3) Dựa trên phương pháp lai (hybrid model)

4.1 Collabrative Filtering

4.1.1 Memory–based CF

Phân thành hai loại: User-based (dựa trên người dùng) và Item-based (dựa trên món ăn).

Sử dụng ma trận đánh giá để thực hiện dự đoán và khuyến nghị. Giả sử mỗi người dùng thuộc ít nhất một nhóm những người có chung sở thích, mối quan tâm. Người cần được khuyến nghị được gọi là active user. Những người dùng có sở thích tương tự với active user được gọi là neighbors. Các bước thực hiện:

Bước 1: Similarity computation - Tính toán độ tương tự, sử dụng tương quan Pearson.

(1) User-based CF: Tính toán độ tương tự ($w_{u,v}$) giữa hai người dùng u, v dựa trên những đánh giá của hai người dùng này trên cùng các items.

$$w_{u,v} = \frac{\sum_{i \in I}(r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I}(r_{u,i} - \bar{r}_u)^2}\sqrt{\sum_{i \in I}(r_{v,i} - \bar{r}_v)^2}} \quad (1)$$

Trong đó:

- I: Tập các món ăn mà cả hai người dùng u và v cùng đánh giá.
- $r_{u,i}$ và $r_{v,i}$: Các ratings mà người dùng u và v đánh giá cho món ăn i.
- $\bar{r}_u$ và $\bar{r}_v$: Trung bình ratings của người dùng u và v trên các món ăn I.

(2) Item-based CF: Tính toán độ tương tự (w_(i,j)) giữa hai item i và j dựa trên những người dùng cùng đánh giá hai món ăn này.

$$w_{i,j} = \frac{\sum_{u \in U}(r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U}(r_{u,i} - \bar{r}_i)^2}\sqrt{\sum_{u \in U}(r_{u,j} - \bar{r}_j)^2}} \quad (2)$$

Trong đó:

- U: Tập các người dùng cùng rating cho hai món ăn i và j.
- $r_{u,i}$ và $r_{u,j}$: Ratings mà người dùng u đánh giá cho các món ăn i và j.
- $\bar{r}_i$ và $\bar{r}_j$: Trung bình rating của các người dùng trong U cho món ăn i và j.

Bước 2: Prediction - Dựa trên thông tin từ bước 1 để dự đoán đánh giá của một người dùng (active user) cho món ăn chưa đánh giá, sử dụng thuật toán KNN và các biến thể của nó KNN-Baseline, KNN-Mean. Dựa trên độ tương đồng, thuật toán sẽ tìm ra k đối tượng gần nhất với đối tượng cần dự đoán. Thuật toán dự đoán sẽ sử dụng điểm đánh giá của các đối tượng gần nhất để dự đoán điểm đánh giá của đối tượng cần dự đoán.

Điểm đánh giá dự đoán được tính như sau:

(1) Thuật toán KNN:

$$est_{knn} = \frac{\sum_{\text{neighbors}}(\text{similarity} \times \text{rating})}{\sum_{\text{neighbors}} \text{similarity}} \quad (3)$$

Trong đó:

- $\sum_{\text{neighbors}}$: Biểu thị việc tổng hợp qua tất cả các hàng xóm trong $K_{\text{neighbors}}$, các hàng xóm này được chọn từ tất cả các người dùng (User-based) hoặc từ tất cả các món ăn (Item-based).
- similarity: Độ đo tương đồng giữa người dùng (User-based) hoặc giữa món ăn (Item-based) vời một hàng xóm cụ thể.
- rating: Là rating thực tế mà hàng xóm đó đã đánh giá cho món ăn i.

(2) Thuật toán KNN Baseline:

$$\text{est} = \text{global\_mean} + b_u[u] + b_i[i] + \sum_{\text{neighbors}} \text{similarity}(i,j) \cdot (r_{uj} - \text{baseline}(i,j)) \quad (4)$$

Trong đó:

- global_mean: Đánh giá trung bình toàn cục của tất cả các đánh giá trong tập dữ liệu.
- $b_u[u]$: Ước lượng độ lệch người dùng (bias user) cho người dùng u.
- $b_i[i]$: Ước lượng độ lệch mục (bias item) cho món ăn i.
- $\sum_{\text{neighbors}}$: Biểu thị việc tổng hợp qua tất cả các hàng xóm trong $k_{\text{neighbors}}$, các hàng xóm này được chọn từ tất cả các người dùng (User-based) hoặc từ tất cả các món ăn (Item-based).
- similarity: Độ đo tương đồng giữa người dùng (User-based) hoặc giữa món ăn (Item-based) và một hàng xóm cụ thể.

- r: Đánh giá thực tế mà hàng xóm đó đã đánh giá cho món ăn i.

- baseline: Ước lượng đánh giá cơ bản cho món ăn i của người dùng hàng xóm, được tính dựa trên trung bình toàn cục, độ lệch người dùng và độ lệch món ăn.

(3) Thuật toán KNN Means:

$$est_{knnmeans} = \mu_u + \sum_{\text{neighbors}} \text{similarity} \cdot (r - \mu_v) \tag{5}$$

Trong đó

- μ_u là trung bình rating của người dùng u.

- μ_v là trung bình rating của hàng xóm v.

- $\sum_{\text{neighbors}}$ Biểu thị việc tổng hợp qua tất cả các hàng xóm trong $k_{\text{neighbors}}$, các hàng xóm này được chọn từ tất cả các người dùng (trong trường hợp lọc cộng tác theo người dùng) hoặc từ tất cả các món ăn (trong trường hợp lọc cộng tác theo mục).

- similarity là độ đo tương đồng giữa người dùng (User-based) hoặc giữa món ăn (Item-based) và một hàng xóm cụ thể.

- r là đánh giá thực tế mà hàng xóm đó đã đánh giá cho món ăn i.

Bước 3: Gợi ý top-N món ăn

(1) Top-N sản phẩm theo người dùng

- Gọi a là active user.

- Gọi U_a là tập k người dùng tương tự nhất với a (được tính ở 2 bước trên).

- Gọi C là tập tất cả các món ăn mà các người dùng trong U_a đã đánh giá mà a chưa đánh giá.

- Xếp hạng các món ăn trong C giảm dần theo số người dùng (trong U_a) đánh giá.

- Lấy top N món ăn từ C theo thứ tự xếp hạng trên để gợi ý cho a.

(2) Top-N sản phẩm theo món ăn

- Gọi a là active user, R là ma trận đánh giá.

- Gọi I_a là tập món ăn mà a đã đánh giá.

- Với mỗi món ăn trong I_a, xác định k món ăn tương tự nhất với i.

- C là tập tất cả các món ăn tương tự các món ăn trong I_a.

- Loại bỏ các món ăn trong I_a khỏi C.

- Tính độ tương tự giữa các món ăn trong C với tập món ăn I_a.

- Xếp hạng C giảm dần theo mức độ tương tự nói trên.

- Lấy top n món ăn từ C theo thứ tự giảm dần của độ tương tự, sau đó gợi ý cho người dùng a.

4.1.2 Model-based CF:

Để đánh giá và xếp hạng tương tác của người dùng với các item mà họ chưa tương tác. Các mô hình này được phát triển thông qua nhiều thuật toán khác nhau, như phân giải ma trận, học sâu, phân cụm, v.v., sử dụng dữ liệu tương tác đã có trong ma trận tương tác. Hơn nữa, một mô hình có thể được sử dụng để đề xuất item và dự đoán đánh giá cho chưa được đánh giá. Phương pháp này giúp giảm bớt chi phí bộ nhớ mà phương pháp dựa trên bộ nhớ mang lại và rất hữu ích khi làm việc với ma trận rải rác. Trong nghiên cứu này, nhóm thực nghiệm bốn mô hình:

- **Singular value decomposition (SVD)**: là một kỹ thuật phân rã ma trận, thường được sử dụng để giảm chiều dữ liệu và khám phá các yếu tố ẩn, chia ma trận ban đầu thành các ma trận phụ chứa các đặc trưng ẩn, đó không phải là các đặc trưng thực sự có mặt trong tập dữ liệu mà thay vào đó là những đặc trưng ẩn có giá trị mà thuật toán phát hiện ra. Nó dự đoán các đánh giá chưa biết trong một ma trận user-item bằng cách sử dụng tích vô hướng của U (ma trận user-user), Σ (ma trận user-item) và V^T (ma trận item-item được chuyển vị). Mặc dù chỉ có thể sử dụng k số đặc trưng để xấp xỉ ma trận ban đầu M, phân giải giá trị đơn đẳng xem xét tất cả $m \times n$ đặc trưng trong tính toán. Đây là khái niệm đằng sau việc giảm chiều của lọc hợp tác dựa trên mô hình. Ngoài ra, giả sử rằng ma trận đặc trưng được gập vào U và V^T khi được sử dụng cho hệ thống đề xuất, chúng ta có thể dự đoán ma trận đánh giá "xấp xỉ" M' thông qua tích vô hướng của U và V^T.

$$M_{m \times n} = U_{m \times m} \Sigma_{m \times n} V^T_{n \times n} \tag{6}$$

- **Co-Clustering (Phân nhóm cộng tác):** Đây là một kỹ thuật phân nhóm cả người dùng và sản phẩm. Mô hình này giả định rằng những người dùng thuộc cùng một nhóm sẽ có xu hướng đánh giá tương tự với các sản phẩm thuộc cùng một nhóm. Việc này giúp tạo ra các nhóm người dùng và sản phẩm có sự tương đồng trong cách họ tương tác và đánh giá.

- **Non-negative Collaborative Filtering:**

- **Neural Collaborative Filtering (Neural CF)** là một phương pháp trong model-based collaborative filtering, trong đó mạng nơ-ron được sử dụng để học các tương tác giữa người dùng và sản phẩm. Neural CF sử dụng kiến trúc mạng nơ-ron để mô hình hóa các mối quan hệ phi tuyến tính giữa các yếu tố người dùng và sản phẩm. Kết hợp giữa GMF (Generalized Matrix Factorization) và MLP (Multi-Layer Perceptron) được kết hợp thông qua một lớp kết hợp hoặc các phép toán khác để tạo ra điểm số dự đoán cuối cùng. GMF sử dụng lớp biểu diễn tuyến tính để học đặc trưng từ userID và itemID, các đặc trưng này sau đó được kết hợp thông qua một phép toán tuyến tính hoặc element-wise để tạo ra đầu ra. MLP là một mạng nơ-ron feedforward với nhiều lớp ẩn, sử dụng các lớp nhúng để biểu diễn userID và itemID và học cách kết hợp chúng qua các lớp ẩn.

4.2 Content-based Filtering

Phương pháp Content-based Filtering (Lọc dựa trên nội dung) là một kỹ thuật được sử dụng trong hệ thống đề xuất để gợi ý sản phẩm hoặc thông tin dựa trên sự phân tích của nội dung liên quan. Trong phần nghiên cứu phương pháp lọc dựa trên nội dung này, nhóm thực thực hiện phát triển hệ thống gợi ý với dữ liệu đầu vào là thành phần và chất dinh dưỡng của món ăn, được thực hiện qua các bước sau:

B1 ***Tiền xử lý dữ liệu:*** Chuyển đổi cột nutrients thành một DataFrame số. Sau đó chuẩn hóa các giá trị dinh dưỡng trong DataFrame này để loại bỏ ảnh hưởng của sự chênh lệch giữa các đặc trưng.

B2 ***Trích xuất đặc trưng:*** Sử dụng TF-IDF để vector hóa thành phần của các món ăn. Sau đó, kết hợp vector đặc trưng của thành phần và chất dinh dưỡng để tạo ra một biểu diễn tổng hợp cho mỗi món ăn.

B3 ***Tính toán độ tương đồng:*** Sử dụng độ tương tự cosine để đo lường mức độ giống nhau giữa các món ăn dựa trên các đặc trưng tổng hợp của chúng.

B4 ***Xây dựng mô hình:*** Sử dụng mô hình hồi quy Ridge để học mối quan hệ giữa các đặc trưng và đánh giá trung bình của các món ăn. Mô hình hồi quy Ridge là một mô hình hồi quy tuyến tính có thêm một thuật toán Regularization để giảm bớt hiện tượng overfitting.

B5 ***Tạo khuyến nghị:*** Top n món ăn có độ tương đồng cao nhất với món ăn mà người dùng quan tâm sẽ được đề xuất cho người dùng như những lựa chọn tiềm năng.

4.3 Hybrid Model Approach - Mô hình LightFM

Trong cách tiếp cận này, chúng tôi giới thiệu một mô hình hybrid sử dụng LightFM, một thuật toán phổ biến trong hệ thống khuyến nghị. Mô hình LightFM hoạt động bằng cách mỗi người dùng và mỗi sản phẩm được biểu diễn bởi một tập hợp các đặc trưng tiềm ẩn, tính năng của mỗi người dùng và sản phẩm được nhúng vào không gian tiềm ẩn.

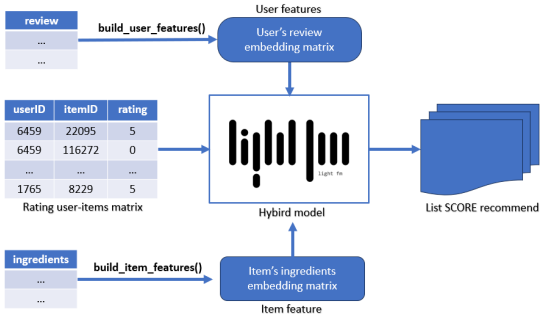

Figure 4: Sơ đồ Kiến trúc Mô hình khuyến Nghị Hybrid sử dụng LightFM

LightFM có khả năng tích hợp cả lọc cộng tác và lọc dựa trên nội dung để tận dụng ưu điểm của cả hai phương pháp với lọc cộng tác tập trung vào các mối quan hệ giữa người dùng và sản phẩm dựa trên hành vi của cộng đồng người dùng và lọc dựa trên nội dung tập trung vào các đặc trưng cụ thể của người dùng và sản phẩm. LightFM có khả năng đối mặt với vấn đề cold-start cho cả người dùng mới và sản phẩm mới bằng cách tích hợp thông tin từ cả hai loại. Trong nghiên cứu này, chúng tôi đưa ra 2 cách tiếp cận cho mô hình LightFM:

- LightFM_1: Ban đầu, mô hình được huấn luyện trên ma trận đánh giá user-item, nơi mỗi hàng biểu diễn một user và mỗi cột đại diện cho một item đặc trưng. Qua đó, mô hình học được xu hướng và sở thích cá nhân của từng người dùng dựa trên lịch sử đánh giá của họ.

- LightFM_2: Tiếp theo, mô hình được mở rộng bằng cách tích hợp thông tin đặc trưng cho cả người dùng và sản phẩm. Với mục tiêu tăng cường độ chính xác của việc khuyến nghị, chúng tôi nhúng thông tin về thành phần nguyên liệu của món ăn như một phần đặc trưng sản phẩm (*4.2*), cùng với việc sử dụng đánh giá (điểm số và nhận xét) từ người dùng như là đặc trưng người dùng. Kết quả là một mô hình đa dạng, có khả năng dự đoán khuyến nghị dựa trên sở thích cụ thể và nhu cầu dinh dưỡng của người dùng.

4.4 Hệ Thống Khuyến Nghị Dựa Trên Phân Tích Cảm Xúc

Trong nghiên cứu này, chúng tôi phát triển một hệ thống khuyến nghị thực phẩm mới, kết hợp giữa lọc nội dung truyền thống và phân tích cảm xúc từ đánh giá người dùng. Hệ thống thực thi theo các bước trình bày trong **Algorithm 1**.

Algorithm 1 Tổng quan phần mã giả

1: Khởi tạo danh sách các món ăn người dùng đánh giá cao.
2: Khởi tạo danh sách các món ăn chưa được đánh giá bởi người dùng.
3: **Vòng lặp:** Cho mỗi món ăn người dùng thích.
4: Tính toán độ tương đồng (*cosine sim*) với các món ăn chưa đánh giá (*4.2*).
5: Tính điểm đánh giá từ phân tích cảm xúc của đánh giá người dùng (*sentiment score*).
6: Kết hợp điểm tương đồng (*cosine sim*), điểm đánh giá (*rating score*) và điểm cảm xúc (*sentiment score*) để tạo điểm tổng hợp (*fusion score*).
7: **Kết thúc vòng lặp.**
8: Kết hợp tất cả các danh sách đã sắp xếp vào danh sách cuối cùng.
9: **Vòng lặp:** Cho mỗi món ăn trong danh sách cuối cùng.
10: Tính điểm cuối cùng (*final score*) bằng cách lấy trung bình cộng các điểm tổng hợp.
11: **Kết thúc vòng lặp.**
12: Sắp xếp danh sách cuối cùng dựa trên điểm cuối cùng.
13: Đề xuất top n các món ăn dựa trên điểm cuối cùng.

Trong phần Thuật toán tính điểm đánh giá từ phân tích cảm xúc của đánh giá người dùng, chúng tôi sử dụng một mô hình phân tích cảm xúc (mô hình LSTM (Long Short-Term Memory) làm thành phần chính), mô hình này được đào tạo trên tập dữ liệu gồm các reviews được gắn nhãn 'tích cực' hoặc 'tiêu cực' (rating của người dùng lớn hơn rating trung bình của món ăn thì đánh nhãn là 'tích cực', và ngược lại). Sau đó, sử dụng điểm phân loại của mô hình làm điểm đánh giá.

Công thức tổng quát để kết hợp 3 yếu tố/điểm số thành một điểm số tổng hợp (fusion score):

$$\text{Fusion Score} = \alpha \cdot S_i + \beta \cdot R_i + \gamma \cdot C_i \quad (7)$$

Trong đó:

- S_i là điểm tương tự từ Content-based Filtering cho món ăn i.
- R_i là điểm đánh giá trung bình cho món ăn i.
- C_i là điểm cảm xúc từ Sentiment Analysis cho món ăn i.
- α, β, γ là các trọng số được gán cho mỗi phần, phản ánh mức độ quan trọng tương đối của Content-based Filtering, Rating, và Sentiment Analysis trong quyết định cuối cùng.

4.5 Các cài đặt thực nghiệm

Trong cách tiếp cận **Collaborative Filtering**, các thông số của các mô hình được mô tả trong bảng 4.

Algorithm	Parameter	Value
KNN	Neighbors (k)	40
	Độ tương tự	Pearson
KNN Baseline	Neighbors (k)	40
	Độ tương tự	Pearson
	Baseline	ALS
KNN Means	Neighbors (k)	40
	Độ tương tự	Pearson
Co-Clustering	Số cụm cho user	3
	Số cụm cho item	3
	Epochs	20
SVD	Factors	100
	Epochs	20
	Learning rate	0.005
Non-negative MF	Factors	15
	Epochs	50
	Learning rate	0.005
Neural CF	Epochs	50, 100
	Batch size	64

Table 4: Thông số các mô hình Collaborative Filtering

Trong cách tiếp cận **Content-based Filtering**

- Sử dụng `TfidfVectorizer` để chuyển đổi thành phần món ăn thành ma trận TF-IDF, sau đó kết hợp dữ liệu dinh dưỡng đã chuẩn hóa.
- Mô hình Ridge Regression được huấn luyện với giá trị alpha là 1.0.

Trong cách tiếp cận **Hybrid - LightFM**: Mô hình được cấu hình với các thông số được đề cập trong bảng 5.

Parameter	Value
Số chiều feature embedding	23
Thuật toán tối ưu	adagrad
Hàm mất mát	warp
Tốc độ học	0.99377
Số mẫu lấy tối đa	12
Epoch	43

Table 5: Thông số huấn luyện mô hình LightFM

Trong phương pháp hệ thống khuyến nghị dựa trên **phân tích cảm xúc**, chúng tôi thiết lập công thức Fusion score với: $\alpha = 0.5$, $\beta = 0.3$, $\gamma = 0.2$.

Tất cả được thực nghiệm trên môi trường Google Colab, với một card NVIDIA T4 GPU duy nhất.

5 Đánh giá

5.1 Thang đo đánh giá

Để phân tích hiệu suất của các thuật toán được đề xuất, các phép đo được chúng tôi sử dụng:

- Root-mean-square deviation(RMSE): Độ lệch trung bình bình phương là căn bậc hai của trung bình cách bình phương giữa giá trị thực và giá trị dự đoán. Nó được sử dụng trong các vấn đề hồi quy để ước lượng tỷ lệ lỗi và xác định xem có bất kỳ lỗi lớn nào do mô hình đánh giá quá cao hoặc đánh giá thấp không. Công thức là Độ lệch Trung bình Bình phương, trong đó sự tổng hợp của các thành phần được thực hiện, và giá trị thực và giá trị dự đoán được trừ và bình phương, sau đó số này được chia cho số điểm dữ liệu. Cuối cùng, kết quả được lấy căn bậc hai.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{8}$$

- Mean Absolute Error (MAE): Sai số trung bình tuyệt đối là một công thức được sử dụng để tính độ lỗi so sánh giữa giá trị dự đoán và giá trị thực của các điểm dữ liệu. Công thức (9) làm rõ rằng công thức này tìm giá trị trung bình của tất cả các sai số tuyệt đối. Ở đây, $\hat{y}_i$ là giá trị dự đoán thứ i, và y_i là giá trị thực.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{9}$$

- Normalized Mean Absolute Error (NMAE) là một phép đo lường sự chênh lệch trung bình tuyệt đối giữa giá trị dự đoán và giá trị thực, được chuẩn hóa theo phạm vi giá trị thực. NMAE thường được sử dụng để đánh giá độ chính xác của mô hình dự đoán trong các vấn đề regression.Giá trị NMAE càng thấp, mô hình càng chính xác. Nó được tính theo công thức (10), trong đó $\hat{y}_i$ là giá trị dự đoán thứ i, y_i là giá trị thực, và $y_{\text{max}} - y_{\text{min}}$ là phạm vi giá trị thực.

$$\text{NMAE} = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{n(y_{\text{max}} - y_{\text{min}})} \tag{10}$$

Precision, Recall và F1-measure là ba trong số các thước đo thường xuyên được sử dụng trong lĩnh vực truy xuất thông tin. Để tính toán các thước đo này, ma trận nhầm lẫn được sử dụng để phân loại các mục thành bốn nhóm. Ma trận này đặt các mục liên quan mà hệ thống đề xuất là liên quan cho người dùng vào hạng mục true positive (TP), và các mục liên quan mà hệ thống không nhận ra là liên quan cho người dùng vào hạng mục false negative (FN). Các mục liên quan được đề xuất sai lầm bởi hệ thống được xem xét là false positives (FP) và các mục không liên quan được đề xuất đúng cho người dùng được xem xét là true negatives (TN). Dựa trên ma trận nhầm lẫn, Precision được định nghĩa là tỷ lệ giữa số mục liên quan được đề xuất và tổng số mục được đề xuất, như sau:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Recall được định nghĩa là tỷ lệ giữa số mục liên quan được đề xuất và tổng số tất cả các mục liên quan. Có một sự xung đột rõ ràng giữa các tiêu chí Precision và Recall. Tăng số lượng các mục được đề xuất hàng đầu sẽ tăng số lượng các mục liên quan và cũng đo lường Recall, trong khi giảm đo lường Precision. Bằng cách kết hợp Precision và Recall, F1 cung cấp một đo lường kết hợp trọng số phù hợp.

Precision và Recall của các thuật toán đề xuất không thể được đánh giá trực tiếp vì chúng ta cần biết liệu mỗi mục có liên quan hay không, điều này có nghĩa là mỗi mục nên được người dùng đánh giá. Do đó, trong các thử nghiệm của chúng tôi, chúng tôi sử dụng Precision@N, Recall@N và F1@N (N là kích thước của danh sách đề xuất). Các đo lường

này có thể được tính toán như sau:

$$\text{Precision@N} = \frac{TP}{N}$$

$$\text{Recall@N} = \frac{TP}{|\text{Relu}|}$$

$$\text{F1@N} = \frac{2 \times \text{Precision@N} \times \text{Recall@N}}{\text{Precision@N} + \text{Recall@N}}$$

Ở đây, |Relu| là số lượng mục liên quan thực tế. Trong đó, |Relu| chỉ số số lượng các mục có liên quan đến người dùng u.

Hit Ratio là một thước đo thường được sử dụng trong lĩnh vực truy xuất thông tin và đề xuất hệ thống để đánh giá hiệu suất của các hệ thống đề xuất. Hit Ratio được tính bằng cách chia số lượng truy vấn mà hệ thống đề xuất ít nhất một mục liên quan cho tổng số truy vấn. Công thức của Hit Ratio được biểu diễn như sau:

$$\text{Hit Ratio} = \frac{\text{off cache hits}}{\text{off cache hits + off cache misses}}$$

Trong trường hợp này, "mục liên quan" có thể được xác định dựa trên đánh giá của người dùng hoặc các tiêu chí đặc biệt của vấn đề cụ thể mà hệ thống đang giải quyết. Hit Ratio giúp đo lường khả năng của hệ thống đề xuất trong việc đưa ra những gợi ý phù hợp với nhu cầu của người dùng.

Normalized Discounted Cumulative Gain (NDCG), có giá trị trong khoảng đơn vị, là một độ đo được sử dụng trong các thử nghiệm của chúng tôi. NDCG gán giá trị cao hơn cho các hit ở vị trí cao trên danh sách xếp hạng. Giá trị NDCG cao chỉ ra rằng danh sách gợi ý có khả năng hoạt động hiệu quả hơn đối với các mục có liên quan.

$$\text{NDCG} = \sum_{i=1}^{N} \frac{2^{rel_i} - 1}{\log_2(i+1)}$$

5.2 Kết quả thực nghiệm

		Metrics Measure		
Type	**Algorithm**	**MAE**	**NMAE**	**RMSE**
Item-based(Memory-based)	KNN-basic	0.524	0.105	0.971
	KNN-w-Baseline	0.510	0.102	0.950
	KNN-w-Means	0.518	0.104	0.957
User-based (Memory-based)	KNN-basic	0.597	0.119	1.024
	KNN-w-Baseline	0.534	0.107	0.966
	KNN-w-	0.520	0.104	0.958
Model-based	Co-Clustering	0.497	0.099	0.899
	SVD	**0.487**	**0.097**	**0.854**
	Non-negative MF	0.624	0.125	0.928
Content-based	Ridge Regression	0.19	0.038	0.257

Table 8: Kết quả phương pháp Collaborative filtering và Content-based Filtering.

	Accuracy
Validation	84.33
Test	84.55

Figure 5: Độ chính xác của mô hình

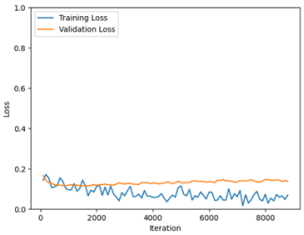

Figure 6: Đường cong học tập của mô hình.

5.3 Phân tích kết quả

Đối với kết quả dựa trên các độ đo MAE, RMSE và NMAE:

Trong thực nghiệm này, thuật toán KNN-w-Baseline có vẻ cho kết quả tốt nhất ở cả hai loại thuật toán, với giá trị MAE và RMSE thấp hơn so với các thuật toán khác.

Ở phần thuật toán dựa trên mô hình, Co-Clustering và SVD có kết quả tương đối tốt với giá trị MAE và RMSE thấp. Đặc biệt, SVD cho thấy sự ổn định và chính xác cao hơn so với các thuật toán khác trong cùng phân khúc.

Ngoài ra, thuật toán Content-based, đặc trưng bởi Ridge Regression, thể hiện kết quả ấn tượng nhất với giá trị MAE, NMAE và RMSE thấp nhất trong tất cả các thuật toán đã được đánh giá. Điều này cho thấy tính chính xác cao và hiệu suất tốt của thuật toán dựa trên nội dung trong việc đưa ra các gợi ý.

Đối với kết quả dựa trên độ đo P@k, R@k, Hit Ratio và NDCG:

- Phương pháp Neural Collaborative Filtering (NCF): có hiệu suất tốt hơn so với Sentiment-Network và Hybrid (LightFM-1, LightFM-2). Epoch = 50, Batch_size = 64 có kết quả tốt nhất với các chỉ số đánh giá như Precision@10, Recall@10, F1@10, Hit Ratio, và NDCG. Kết quả của Epoch = 100 không có sự cải thiện đáng kể so với Epoch = 50.

- Phương pháp SentimentNetwork và Hybrid: SentimentNetwork có hiệu suất thấp hơn rất nhiều so với NCF ở mọi mức độ đánh giá. Hybrid (LightFM-1 và LightFM-2) cũng cho kết quả tương đối ổn định.

Tổng quan, Neural Collaborative Filtering (NCF) là phương pháp gợi ý tối ưu nhất. NCF này đã đạt được hiệu suất tốt nhất dựa trên các chỉ số đánh giá chất lượng của hệ thống gợi ý, vượt trội hơn so với các phương pháp dựa trên bộ nhớ truyền thống.

Algorithm	Epoch & Batch_size	With pre-train					Without pre-train				
		Precision@10	Recall@10	F1@10	Hit Ratio	NDCG	Precision@10	Recall@10	F1@10	Hit Ratio	NDCG
Neural CF	Epoch = 50, Batch_size = 64	0.017005	0.017226	0.017115	**0.274061**	0.19717	0.014372	0.015207	0.014778	**0.24126**	0.178988
Neural CF	Epoch = 100, Batch_size = 64	0.016444	0.016821	0.016630	**0.287872**	0.206201	0.014847	0.015268	0.015055	**0.245145**	0.175297

Table 6: Kết quả phương pháp Neural Collaborative Filtering (NCF)

Algorithm	Precision@10	Recall@10	F1@10	Hit Ratio	NDCG
SentimentNetwork	0.004434	0.004619	0.004002	0.040465	0.005429
LightFM-1	0.013233	0.014293	0.013745	-	-
LightFM-2	**0.013379**	**0.014389**	**0.013865**	-	-

Table 7: Kết quả các phương pháp Hybrid

Tuy nhiên, Hybrid cũng có tiềm năng khi kết hợp các phương pháp khác nhau giúp cải thiện hiểu biết về khả năng của nó trong hệ thống gợi ý.

6 Kết luận

Trong nghiên cứu này, chúng tôi đã áp dụng các phương pháp gợi ý thực phẩm, bao gồm tìm kiếm công thức và gợi ý thực phẩm dựa trên thông tin sức khỏe. Để chứng minh tính hiệu quả của phương pháp kết hợp giữa truyền thống và hiện đại của chúng tôi, chúng tôi xây dựng một bộ dữ liệu chất lượng cao. Kết quả thực nghiệm đã chứng minh sự hiệu quả của phương pháp gợi ý thực phẩm Neural Collaborative Filtering và sự vượt trội trong hiệu suất của các mô hình được đề xuất. Hơn nữa, từ kết quả thực nghiệm, chúng tôi rút ra các kết luận sau: Các phương pháp sử dụng thuật toán đạt được chất lượng vượt trội so với phương pháp dựa trên bộ nhớ truyền thống. Tuy nhiên, phương pháp lai cũng có tiềm năng khi kết hợp với các phương pháp khác, giúp cải thiện hiểu biết về khả năng của nó trong hệ thống gợi ý.

Trong tương lai, chúng tôi dự kiến sẽ tập trung vào nâng cao khả năng đối mặt với độ phức tạp của dữ liệu và tăng cường sức mạnh hiểu biết của hệ thống. Nghiên cứu sẽ mở rộng phạm vi đề xuất từ đồ ăn sang các yếu tố như khẩu vị cá nhân, yêu cầu dinh dưỡng, hay sự ưa thích đặc biệt của người dùng. Đồng thời, việc tích hợp các phương pháp tiên tiến như BERTology và các mô hình đọc hiểu máy học tiếng Việt sẽ là những hướng phát triển quan trọng để cải thiện độ chính xác và độ tin cậy của hệ thống.

Acknowledgements

Chúng tôi xin bày tỏ lòng biết ơn đến Giảng viên Huỳnh Văn Tín, những ý kiến của Thầy đã đóng góp quan trọng vào việc nâng cao chất lượng bài báo của chúng tôi. Trong quá trình làm việc, do khả năng của nhóm còn hạn chế, nên có thể xảy ra một số sai sót, chúng tôi sẽ tiếp tục cải thiện để đạt được sự hoàn thiện nhất có thể.

References

[1] Yap, G.-E., Li, X.-L., & Yu, P. S. (2012). *Effective Next-Items Recommendation via Personalized Sequential Pattern Mining*. Database Systems for Advanced Applications, 48–64. https://doi.org/10.1007/978-3-642-29035-0_4

[2] Forouzandeh, Saman, et al. "Health-aware food recommendation system with dual attention in heterogeneous graphs." *Computers in Biology and Medicine (2023)*, 2023.

[3] Cruz, Z. M. C., Alpay, J. J. R., Depeno, J. D. D., Altabirano, M. J. C., and Bringula, R. "Usability of 'Fatchum': A mobile application recipe recommender system." *Proceedings of the 6th Annual Conference on Research in Information Technology*, Mexican International Conference on Artificial Intelligence, 11-16, 2017.

[4] Bowman, Samuel, Angeli, Gabor, Potts, Christopher, and Manning, Christopher D. "A large annotated corpus for learning natural language inference." *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 632–642, 2015.

[5] Ribeiro, D., Ribeiro, J., Vasconcelos, M. J. M., Vieira, E. F., and Barros, A. C. D. "Improved meal recommender system for Portuguese older adults." *Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and e-Health*, Porto, Portugal, pages 107-126, 2017.

[6] Tran, T. N. Trang, Atas, M., Felfernig, A., and Stettinger, M. "An overview of recommender systems in the healthy food domain." *Journal of Intelligent Information Systems*, pages 501–526, 2018.

[7] Yap, G.-E., Li, X.-L., and Yu, P. S. "Effective next-items recommendation via personalized sequential

pattern mining." *Proceedings of the International Conference on Database Systems for Advanced Applications*, pages 48-64, 2012.

[8] Rostami, Mehrdad, et al. "A novel healthy and time-aware food recommender system using attributed community detection." *Expert Systems with Applications* 221 (2023): 119719.

[9] Mahajan, Ketan, et al. "Restaurant Recommendation System using Machine Learning." *International Journal* 10.3 (2021).

[10] Hietala, Joona. "Matrix factorization algorithms for personalized product recommendation: a case study." (2021).

[11] Gu, Yile, and Qingyi Chen. "Predicting Users' Ratings on Yelp based on Various Recommender System Implementations."

[12] Zhou, Zhenxiang, and Lan Xu. "Amazon Food review classification using deep learning and recommender system." *Stanford University Stanford*, 2009.

[13] Chen, Chih-Han, and Christofer Toumazou. "Personalized expert recommendation systems for optimized nutrition." In *Trends in Personalized Nutrition* (2019): 309-338.