

MỘT PHƯƠNG PHÁP LỌC CỘNG TÁC THEO NGỮ CẢNH

Đỗ Thị Liên

Học viện Công nghệ Bưu chính Viễn thông

Tóm tắt: Lọc cộng tác là hướng tiếp cận điển hình trong xây dựng hệ tư vấn. Về cơ bản các phương pháp lọc cộng tác theo cách tiếp cận truyền thống chủ yếu tập trung khai thác duy nhất tập dữ liệu đánh giá của người dùng đối với sản phẩm để thực hiện huấn luyện và dự đoán sản phẩm phù hợp với người dùng hiện thời. Tuy nhiên trên thực tế đánh giá của người dùng với sản phẩm thường thay đổi theo ngữ cảnh khác nhau, do đó hướng tiếp cận lọc cộng tác dựa vào ngữ cảnh đang là một hướng đi được chứng minh là cải thiện đáng kể chất lượng gợi ý trong nhiều ứng dụng thực tế so với hướng tiếp cận lọc cộng tác truyền thống. Đã có nhiều phương pháp đề xuất khác nhau được đưa ra để giải quyết bài toán lọc cộng tác theo ngữ cảnh, tuy nhiên lĩnh vực này vẫn còn tồn tại nhiều thách thức, trong đó khó khăn điển hình là làm thế nào để tích hợp hiệu quả thông tin ngữ cảnh và giải quyết vấn đề dữ liệu thưa thớt dữ liệu đánh giá. Trong bài báo này, tác giả đề xuất một phương pháp lọc cộng tác theo ngữ cảnh với mong muốn giải quyết hiệu quả việc tích hợp đầy đủ thông tin ngữ cảnh và vấn đề dữ liệu thưa. Trong đó, việc tích hợp ngữ cảnh được thực hiện bằng thủ tục phân tách người dùng theo ngữ cảnh và vấn đề dữ liệu thưa được giải quyết qua việc khai thác đầy đủ mối quan hệ trực tiếp và gián tiếp giữa các người dùng và sản phẩm trong quá trình huấn luyện dựa trên mô hình đồ thị. Kết quả thực nghiệm trên một số bộ dữ liệu thực cho thấy phương pháp đề xuất cải thiện đáng kể chất lượng dự đoán so với các phương pháp tư vấn dựa vào ngữ cảnh cơ sở trước đây.

Từ khóa: Lọc cộng tác dựa vào ngữ cảnh (Context-aware collaborative filtering - CACF); Ngữ cảnh (Context); Hệ tư vấn dựa vào ngữ cảnh (Context-aware recommender system - CARS); Lọc trước theo ngữ cảnh (Contextual pre-filtering); Phân tách người dùng theo ngữ cảnh (User splitting)

I. MỞ ĐẦU

Đứng trước vấn đề quá tải thông tin trên Internet thì cần thiết phải có công cụ hỗ trợ người dùng lựa chọn thông tin và đưa ra quyết định một cách nhanh chóng. Hệ tư vấn ra đời đã giải quyết hiệu quả vấn đề này và mang lại thành công vang dội trong nhiều lĩnh vực, đây là lĩnh vực đang có sự phát triển rất nhanh chóng. Có ba hướng tiếp cận chính để giải quyết bài toán tư vấn là lọc cộng tác, lọc theo nội dung và lọc kết hợp. Trong các cách tiếp cận này thì lọc cộng tác được đánh giá là đơn giản trong cài đặt và có thể lọc được mọi loại thông tin, đặc biệt đối với thông tin được thể hiện dưới dạng dữ liệu đa phương tiện.

Trước đây các nghiên cứu về lọc cộng tác cho hệ tư vấn theo cách tiếp cận truyền thống chỉ tập trung khai thác hai nhóm đối tượng chính là người dùng (user) và sản phẩm (item) để thực hiện huấn luyện và đưa ra gợi ý. Bài toán lọc cộng tác cho hệ tư vấn theo cách tiếp cận truyền thống có thể được biểu diễn dựa trên ma trận đánh giá hai chiều sau:

$$R_0: Users \times Items \rightarrow Ratings \quad (1)$$

Tuy nhiên trên thực tế, quyết định của người dùng thường thay đổi tùy thuộc vào ngữ cảnh. Ví dụ một người có thể chọn xem những bộ phim khác nhau khi xem chung với bạn đời và trẻ em. Hay việc chọn nhà hàng ăn uống là sang trọng hay ăn nhanh cũng tùy thuộc vào đối tượng ta đi ăn cùng là đối tác trong công việc hay đi ăn một mình. Do vậy việc xem xét kết hợp ngữ cảnh vào các hệ thống tư vấn là một chủ đề đang rất được quan tâm nghiên cứu trong những năm gần đây. Ví dụ hệ tư vấn phim và âm nhạc [3], tư vấn địa điểm yêu thích [4], thương mại điện tử [5],... Theo như [2]: “Thông tin ngữ cảnh là những thông tin có thể mô tả được hoàn cảnh của một thực thể. Thực thể ở đây có thể là người, là vật hoặc là đối tượng có liên quan tới sự tương tác giữa người dùng và ứng dụng, bao gồm cả bản thân người dùng và ứng dụng đó”. Chẳng hạn đối với hệ tư vấn du lịch, yếu tố ngữ cảnh có thể là thời gian (buổi trong ngày, thời gian trong tuần, mùa), bạn đồng hành (một mình, gia đình, bạn bè). Bài toán lọc cộng tác theo ngữ cảnh sẽ được biểu diễn dựa trên ma trận đánh giá đa chiều như sau:

$$R_1: Users \times Items \times Contexts \rightarrow Ratings \quad (2)$$

Giả sử ta có tập hữu hạn $U = \{u_1, u_2, \dots, u_N\}$ là tập gồm N người dùng, $P = \{p_1, p_2, \dots, p_M\}$ là tập gồm M sản phẩm và K chiều ngữ cảnh C_1, C_2, \dots, C_K , mỗi chiều ngữ cảnh có tương ứng $N_{C_1}, N_{C_2}, \dots, N_{C_K}$ điều kiện ngữ cảnh. Nhiệm vụ của lọc cộng tác theo ngữ cảnh là dự đoán đánh giá và đưa ra tư vấn các sản phẩm mới cho người dùng hiện thời trong từng tình huống ngữ cảnh cụ thể.

Bảng 1 là một ví dụ về ma trận đánh giá đa chiều của hệ tư vấn cộng tác theo ngữ cảnh, gồm 2 người dùng $U = \{u_1, u_2\}$, 3 sản phẩm $P = \{p_1, p_2, p_3\}$, kèm thông tin về các chiều ngữ cảnh như sau:

Bảng 1. Ma trận đánh giá đa chiều của lọc cộng tác theo ngữ cảnh

U	P	R	Time	Location	Companion
u_1	p_1	5	Weekend	Home	Kids
u_1	p_1	4	Weekday	Home	Family
u_1	p_2	3	Weekend	Cinema	Partner
u_2	p_2	4	Weekday	Home	Family
u_2	p_2	3	Weekend	Cinema	Partner
u_2	p_3	2	Weekend	Cinema	Partner
u_1	p_2	1	Weekday	Home	Family
u_2	p_3	3	Weekday	Home	Family
u_1	p_3	2	Weekday	Home	Family

Tác giả liên hệ: Đỗ Thị Liên,

Email: liendt@ptit.edu.vn

Đến tòa soạn: 03/10/2022, chỉnh sửa: 24/11/2022, chấp nhận đăng: 11/12/2022.

Một số thuật ngữ qui ước được sử dụng khi khai thác ma trận đánh giá đa chiều ở trên là: *Chiều ngữ cảnh (Context Dimension)*, *điều kiện ngữ cảnh (Context Condition)*, *tình huống ngữ cảnh (Context Situation)*. *Chiều ngữ cảnh* hay còn được biết đến là các biến ngữ cảnh, ví dụ: “Time”, “Location”, “Companion”. *Điều kiện ngữ cảnh* là giá trị của một chiều ngữ cảnh, ví dụ: chiều ngữ cảnh “Location” có 2 điều kiện ngữ cảnh (“Home”, “Cinema”). Thuật ngữ *tình huống ngữ cảnh* chỉ một bộ giá trị nhận được cho các điều kiện ngữ cảnh ứng với các chiều ngữ cảnh, ví dụ người dùng u_1 đánh giá sản phẩm p_1 trong tình huống ngữ cảnh (“Weekend”, “Home”, “Kids”) là 5.

Mặc dù đã có một số đề xuất được đưa ra để giải quyết bài toán lọc cộng tác theo ngữ cảnh cho hệ tư vấn, nhưng làm thế nào để tích hợp hiệu quả thông tin ngữ cảnh và giải quyết vấn đề dữ liệu thưa vẫn là vấn đề nghiên cứu mở, có tính thời sự và thu hút được nhiều quan tâm của cộng đồng nghiên cứu. Trong bài báo này, tác giả đề xuất một phương pháp lọc cộng tác theo ngữ cảnh mới cho hệ tư vấn nhằm giải quyết hạn chế còn tồn tại phổ biến, đó là tích hợp đầy đủ thông tin ngữ cảnh và giải quyết hiệu quả vấn đề dữ liệu thưa. Cụ thể, tác giả đề xuất sử dụng phương pháp phân tách người dùng theo ngữ cảnh cải tiến nhằm khắc phục hạn chế nêu trên của phương pháp phân tách người dùng theo ngữ cảnh nguyên thủy, đó là cho phép tích hợp đầy đủ thông tin ngữ cảnh trong việc chuyển hóa người dùng ban đầu thành người dùng giả lập. Tuy nhiên khi áp dụng thủ tục phân tách người dùng theo ngữ cảnh cải tiến lên ma trận đánh giá đa chiều R_1 , với việc giới thiệu các người dùng giả lập, sẽ càng khiến ma trận đánh giá hai chiều R_o thu được càng thưa thớt hơn nữa, điều này sẽ được giải quyết qua việc biểu diễn và xây dựng thuật toán dự đoán dựa trên mô hình đồ thị nhằm khai thác đầy đủ mối quan hệ trực tiếp và gián tiếp giữa các người dùng và sản phẩm. Kết quả thực nghiệm trên một số bộ dữ liệu thực cho thấy phương pháp đề xuất cải thiện đáng kể chất lượng dự đoán so với các phương pháp lọc cộng tác theo ngữ cảnh cơ sở trước đây.

Đề trọng tâm vào phương pháp đề xuất, Mục II tác giả trình bày các nghiên cứu liên quan. Tiếp đến là phương pháp đề xuất trong Mục III. Mục IV trình bày phương pháp thực nghiệm và đánh giá. Mục V nêu kết luận và hướng phát triển trong thời gian tới.

II. CÁC NGHIÊN CỨU LIÊN QUAN

Việc khai thác thông tin ngữ cảnh trong hệ tư vấn được đề cập đầu tiên trong nghiên cứu của Adomavicius and Tuzhilin [1], tiếp sau đó có khá nhiều công trình nghiên cứu liên quan khác được công bố nhằm giải quyết các vấn đề khác nhau của hệ tư vấn dựa vào ngữ cảnh, trong đó tập trung vào cải tiến các phương pháp tư vấn theo ngữ cảnh. Về cơ bản, các phương pháp đề xuất khác nhau để giải quyết bài toán tư vấn cộng tác theo ngữ cảnh thuộc một trong ba hướng tiếp cận: Lọc trước theo ngữ cảnh (Contextual Prefiltering), lọc sau theo ngữ cảnh (Contextual Postfiltering) và mô hình hóa ngữ cảnh (Contextual Modeling) [2][6].

Về cơ bản, các phương pháp lọc cộng tác theo ngữ cảnh thuộc hướng lọc trước ngữ cảnh sử dụng thông tin ngữ cảnh để lọc tập dữ liệu ban đầu nhằm chỉ giữ lại những dữ liệu phù hợp với ngữ cảnh yêu cầu, dữ liệu lọc được sẽ được sử

dùng để huấn luyện và dự đoán bởi các phương pháp lọc cộng tác truyền thống [2][6][8][11]. Một số phương pháp lọc trước ngữ cảnh được đưa ra như: Phương pháp phân chia dữ liệu (Splitting) [26], phương pháp sử dụng một tập con các chiều ngữ cảnh để lọc (Context Relaxation) [27], phương pháp lọc dựa trên ngữ nghĩa (Semantic Filtering) [28]. Tập dữ liệu lọc được sẽ dùng để huấn luyện và tư vấn. Quá trình huấn luyện và tư vấn ở đây có thể sử dụng trực tiếp những phương pháp lọc cộng tác đã được áp dụng cho các hệ tư vấn cộng tác truyền thống để áp dụng cho hệ tư vấn cộng tác theo ngữ cảnh. Ví dụ như một số phương pháp lọc cộng tác như UserKNN, ItemKNN, Matrix Factorization, SLIM, sẽ được áp dụng trực tiếp sau bước lọc trước ngữ cảnh để sinh những sản phẩm dự đoán cho người dùng trong một tình huống ngữ cảnh cụ thể.

Trái ngược với hướng lọc trước ngữ cảnh, lọc sau ngữ cảnh sử dụng toàn bộ ma trận đánh giá đã loại bỏ đi các chiều ngữ cảnh để huấn luyện và tư vấn. Kết quả dự đoán thu được sẽ được lọc lại một lần nữa bởi ngữ cảnh hiện thời để thu được kết quả tư vấn cuối cùng là những sản phẩm mới chưa được người dùng đánh giá trong một tình huống ngữ cảnh cụ thể. Như vậy các phương pháp tư vấn cộng tác theo ngữ cảnh thuộc hướng lọc sau ngữ cảnh cũng có thể áp dụng các phương pháp tư vấn cộng tác truyền thống như lọc trước theo ngữ cảnh. Đối với các phương pháp lọc sau theo ngữ cảnh thì thông tin ngữ cảnh không được tích hợp trong quá trình huấn luyện.

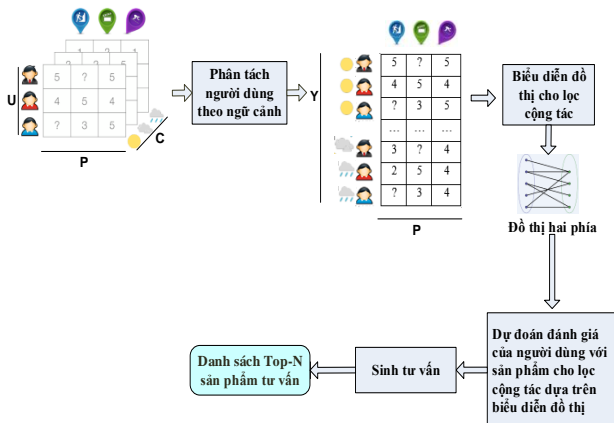
Hướng thứ ba để giải quyết bài toán tư vấn cộng tác theo ngữ cảnh là mô hình hóa dựa vào ngữ cảnh. Theo hướng tiếp cận này thì thông tin ngữ cảnh, người dùng và sản phẩm được biểu diễn trực tiếp trong cùng một mô hình, khi đó ma trận đánh giá đa chiều sẽ được sử dụng trực tiếp cho quá trình huấn luyện và tư vấn. Với hướng tiếp cận này, một số thuật toán lọc cộng tác cho hệ tư vấn cộng tác theo ngữ cảnh được đưa ra có độ phức tạp hơn các phương pháp lọc cộng tác truyền thống. Các phương pháp mô hình hóa dựa vào ngữ cảnh được phân chia thành hai nhóm: Mô hình hóa ngữ cảnh độc lập [12] và mô hình hóa ngữ cảnh phụ thuộc [13][14]. Tensor Decomposition [12] là một phương pháp điển hình thuộc nhóm phương pháp mô hình hóa ngữ cảnh độc lập cho phép biểu diễn người dùng, sản phẩm và các chiều ngữ cảnh trong một không gian đa chiều, mỗi chiều là độc lập nhau. Một vấn đề xảy ra với phương pháp này là không gian lưu trữ cùng xử lý sẽ rất phức tạp khi số lượng chiều ngữ cảnh quá lớn. Không giống như phương pháp mô hình hóa ngữ cảnh độc lập coi ngữ cảnh không phụ thuộc vào người dùng và sản phẩm, phương pháp mô hình hóa ngữ cảnh phụ thuộc sẽ mô tả và khai thác sự phụ thuộc giữa người dùng, sản phẩm và ngữ cảnh tương ứng. Thực nghiệm cho thấy các phương pháp mô hình hóa ngữ cảnh phụ thuộc cho kết quả tốt hơn phương pháp mô hình hóa ngữ cảnh độc lập [7]. Tuy nhiên vấn đề đặt ra với các phương pháp mô hình hóa ngữ cảnh phụ thuộc là khi tích hợp ngữ cảnh vào hệ tư vấn là vấn đề dữ liệu thưa, độ phức tạp tính toán và khả năng mở rộng của nó. Ngoài ra các phương pháp mô hình hóa ngữ cảnh cũng được đánh giá là có độ phức tạp lớn hơn các phương pháp thuộc hướng lọc trước và sau theo ngữ cảnh, đặc biệt khi số chiều ngữ cảnh tăng lên.

Trong bài báo này, tác giả tiếp cận nghiên cứu đề xuất phương pháp tư vấn theo ngữ cảnh mới, cụ thể là một phương pháp tư vấn thuộc hướng lọc trước theo ngữ cảnh. Trong đề xuất của mình, tác giả tiếp cận kết hợp thủ tục phân tách người dùng theo ngữ cảnh cải tiến với dự đoán đánh giá dựa vào mô hình đồ thị cho lọc cộng tác. Khác với

việc sử dụng một chiều ngữ cảnh để phân tách người dùng ban đầu thành các người dùng giả lập không cho phép khai thác triệt để thông tin từ các chiều ngữ cảnh còn lại, thủ tục phân tách người dùng theo ngữ cảnh cải tiến sẽ kết hợp sử dụng tất cả các chiều ngữ cảnh để phân tách người dùng ban đầu thành các người dùng giả lập, điều này giúp tích hợp đầy đủ thông tin ngữ cảnh vào quá trình tư vấn. Việc giới thiệu các người dùng giả lập cũng đồng thời chuyển hóa ma trận đánh giá đa chiều thành ma trận đánh giá hai chiều. Tuy nhiên, bản thân ma trận đánh giá đa chiều ban đầu khá thưa khi được chuyển hóa thành ma trận đánh giá hai chiều càng trở nên thưa thớt hơn nữa. Để giải quyết vấn đề này, tác giả tiếp cận mô hình đồ thị cho phép biểu diễn đầy đủ mối quan hệ trực tiếp và gián tiếp giữa các người dùng và sản phẩm, từ đó huấn luyện và đưa ra dự đoán sản phẩm phù hợp với người dùng.

III. PHƯƠNG PHÁP ĐỀ XUẤT

Về cơ bản phương pháp đề xuất được thực hiện bằng cách kết hợp hai phương pháp: 1) Lọc trước theo ngữ cảnh; 2) Phương pháp dự đoán đánh giá của người dùng với sản phẩm dựa trên mô hình đồ thị đề xuất. Sự kết hợp của hai phương pháp này trong phương pháp đề xuất được thể hiện qua bốn bước: 1) Phân tách người dùng theo ngữ cảnh; 2) Biểu diễn đồ thị cho lọc cộng tác; 3) Dự đoán đánh giá của người dùng với sản phẩm cho lọc cộng tác dựa trên biểu diễn đồ thị; 4) Sinh tư vấn. Bốn bước này kết hợp với nhau trong một bộ khung đề xuất về triển khai phương pháp lọc trước theo ngữ cảnh dựa trên mô hình đồ thị cho hệ tư vấn theo ngữ cảnh trong Hình 1 dưới đây.



Hình 1. Bộ khung triển khai phương pháp lọc trước theo ngữ cảnh dựa trên mô hình đồ thị cho hệ tư vấn cộng tác theo ngữ cảnh

A. Phân tách người dùng theo ngữ cảnh

Thông tin đầu vào cho bài toán lọc cộng tác theo ngữ cảnh gồm có: Tập hợp hữu hạn gồm N người dùng $U = \{u_1, u_1, \dots, u_N\}$, M sản phẩm $P = \{p_1, p_2, \dots, p_M\}$ và K chiều ngữ cảnh C_1, C_2, \dots, C_K , mỗi chiều ngữ cảnh có tương ứng $N_{C_1}, N_{C_2}, \dots, N_{C_K}$ điều kiện ngữ cảnh. Từ thông tin đầu vào trên, việc phân tách người dùng theo ngữ cảnh sẽ chuyển hóa người dùng ban đầu theo ngữ cảnh thành các người dùng giả lập. Mỗi người dùng giả lập được tạo ra từ sự kết hợp người dùng ban đầu với một tình huống ngữ cảnh cụ thể, thủ tục này gọi là “User Splitting” [7].

Thủ tục “User Splitting” trước đây chỉ sử dụng một chiều ngữ cảnh duy nhất để phân tách người dùng theo ngữ cảnh, việc chọn chiều ngữ cảnh này dựa vào độ đo thông kê, như độ lợi thông tin (Information gain) [3][11]. Điều này trong nhiều trường hợp khiến cho khá nhiều thông tin

từ các chiều ngữ cảnh khác không được khai thác triệt để vào quá trình tư vấn sau này.

Từ lập luận đưa ra ở trên, tác giả đề xuất phương pháp phân tách người dùng theo ngữ cảnh cải tiến nhằm khắc phục hạn chế nêu trên của phương pháp phân tách người dùng theo ngữ cảnh nguyên thủy. Phương pháp phân tách người dùng theo ngữ cảnh cải tiến cho phép tích hợp đầy đủ thông tin ngữ cảnh trong việc chuyển hóa người dùng ban đầu thành người dùng giả lập. Các bước thực hiện cụ thể như sau:

- **Bước 1.** Tạo ra 1 chiều ngữ cảnh mới C đại diện cho K chiều ngữ cảnh C_1, C_2, \dots, C_K bằng cách lấy tích Đề-các của tất cả các chiều ngữ cảnh. Khi đó, mỗi điều kiện ngữ cảnh của C là sự kết hợp các điều kiện ngữ cảnh của các chiều tương ứng. Số lượng điều kiện ngữ cảnh của các chiều tương ứng. Số lượng điều kiện ngữ cảnh của C là N_c , với $N_c = N_{c_1} * N_{c_2} * \dots * N_{c_K}$.
- **Bước 2.** Tạo ra tập người dùng giả lập Y bằng cách lấy tích Đề-các của tập người dùng U và chiều ngữ cảnh C . Khi đó, mỗi người dùng giả lập thuộc Y là sự kết hợp của một người dùng ban đầu thuộc U với một điều kiện ngữ cảnh thuộc C . Số lượng người dùng trong tập Y là H , với $H = N * N_c$.
- **Bước 3.** Chuyển đổi ma trận đánh giá đa chiều về ma trận đánh giá hai chiều bằng việc loại bỏ đi tập ngữ cảnh, thay tập người dùng ban đầu U bằng tập người dùng giả lập Y .

Việc thực hiện lấy tích Đề-các giữa các chiều ngữ cảnh và tập người dùng ban đầu cho phép kết hợp đầy đủ các điều kiện ngữ cảnh tương ứng của các chiều ngữ cảnh và các người dùng trong hệ thống. Như vậy tập người dùng giả lập sinh ra sẽ thay thế tương ứng như khi thể hiện dưới dạng tập người dùng ban đầu và các chiều ngữ cảnh gốc. Do đó việc chuyển đổi ma trận đánh giá đa chiều về ma trận đánh giá hai chiều bằng việc loại bỏ đi tập ngữ cảnh, thay tập người dùng ban đầu U bằng tập người dùng giả lập Y sẽ vẫn đảm bảo các điều kiện của bài toán không thay đổi.

Ví dụ áp dụng phương pháp phân tách người dùng theo ngữ cảnh lên ma trận đánh giá đa chiều của lọc cộng tác theo ngữ cảnh (Bảng 1) ta thu được ma trận đánh giá hai chiều (Bảng 2), với y_1 là người dùng giả lập được tạo ra bởi sự kết hợp của người dùng u_1 và tình huống ngữ cảnh (“Cuối tuần”, “Tại nhà”, “Trẻ em”). Với ví dụ được đưa ra trong Bảng 1, hệ tư vấn có 2 người dùng và 12 tình huống ngữ cảnh có thể có, do vậy số lượng người dùng giả lập được sinh ra theo phương pháp phân tách người dùng theo ngữ cảnh cải tiến là 24. Ma trận đánh giá hai chiều nhận được thể hiện trong Bảng 2, tác giả sắp xếp những cặp người dùng - sản phẩm có đánh giá trong những dòng trên cùng của ma trận và những cặp còn lại không có đánh giá ở bên dưới. Để tiết kiệm không gian trình bày, những cặp người dùng - sản phẩm không có đánh giá không nêu đầy đủ trong Bảng 2.

Bảng 2. Ma trận đánh giá hai chiều nhận được sau phân tách người dùng theo ngữ cảnh

Y	P	R
y_1	p_1	5
y_2	p_1	4
y_2	p_3	2

y_3	p_2	3
y_4	p_2	4
y_5	p_2	3
y_6	p_3	2
y_3	p_2	1
y_4	p_3	3
...
y_{24}	p_3	0

Quá trình phân tách người dùng theo ngữ cảnh sẽ biến đổi ma trận đánh giá đa chiều R_1 (biểu diễn đánh giá của người dùng với sản phẩm trong các tình huống ngữ cảnh khác nhau) về ma trận đánh giá hai chiều R_o (biểu diễn đánh giá của người dùng giả lập với sản phẩm). Trên thực tế, số lượng các đánh giá ban đầu đưa ra bởi người dùng cho các sản phẩm trong các tình huống ngữ cảnh là rất ít, khiến cho ma trận R_1 rất thưa. Khi áp dụng thủ tục phân tách người dùng theo ngữ cảnh cải tiến lên R_1 , với việc giới thiệu các người dùng giả lập, sẽ càng khiến ma trận R_o thu được càng thưa thớt hơn nữa.

Để hạn chế những vấn đề dữ liệu thưa của lọc cộng tác áp dụng cho ma trận đánh giá hai chiều R_o , tác giả tiến hành biểu diễn ma trận đánh giá dưới dạng đồ thị hai phía, một phía là người dùng và phía còn lại là sản phẩm. Lựa chọn biểu diễn đồ thị nhằm thể hiện rõ ràng mối quan hệ trực tiếp và gián tiếp giữa người dùng giả lập và sản phẩm của hệ thống, đó là cơ sở để thực hiện đưa ra phương pháp dự đoán đánh giá của người dùng với sản phẩm. Hai nội dung này được thể hiện qua bước 2 và bước 3 trong bộ khung đề xuất và được trình bày trong mục B, C sau đây.

B. Phương pháp biểu diễn đồ thị cho hệ tư vấn cộng tác theo ngữ cảnh

Từ ma trận đánh giá hai chiều nhận được sau phân tách người dùng theo ngữ cảnh, ký hiệu $R_1=(r_{ix})$. Không hạn chế tính tổng quát của bài toán, ta giả sử $r_{ix} = v$ nếu người dùng $y_i \in Y$ đánh giá sản phẩm $p_x \in P$ với mức độ v , trong đó $v \in [0, 1]$.

Để thuận tiện trong trình bày, ta viết $y_i \in U$ ngắn gọn thành $i \in Y, p_x \in P$ ngắn gọn thành $x \in P$.

$$r_{ix} = \begin{cases} v & \text{Nếu người dùng } i \text{ thích sản phẩm } x \text{ ở} \\ & \text{mức độ } v (0 \leq v \leq 1). \\ \phi & \text{Nếu người dùng } i \text{ chưa biết hoặc} \\ & \text{chưa đánh giá sản phẩm } x. \end{cases} \quad (3)$$

Biểu diễn ma trận đánh giá theo (3) sẽ không ảnh hưởng đến các hệ thống tư vấn cộng tác sử dụng đánh giá nhị phân (0,1) hoặc có nhiều mức đánh giá trong khoảng [0,1]. Đối với các bộ dữ liệu có giá trị đánh giá $r_{ix} \in \{1, 2, \dots, V\}$, ta chỉ cần thực hiện phép biến đổi đơn giản chuyển

$$r_{ix} = \frac{r_{ix}}{V}$$

đánh giá theo thứ tự khác nhau của các hệ lọc cộng tác. Đây là một biểu diễn mở rộng của Huang đã thực hiện trong [22].

Ví dụ chuyển đổi ma trận đánh giá đa hai chiều ở Bảng 2 về dạng Bảng 3 như sau:

Bảng 3. Ma trận đánh giá hai chiều chuyển đổi nhận được sau phân tách người dùng theo ngữ cảnh

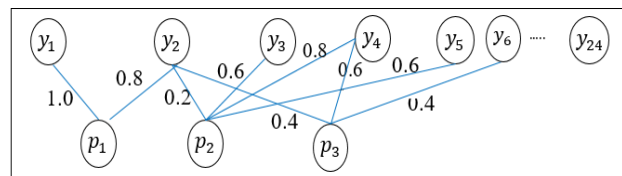
	p_1	p_2	p_3
y_1	1.0	0	0
y_2	0.8	0	0.4
y_3	0	0.6	0
y_4	0	0.8	0.6
y_5	0	0.6	0
y_6	0	0	0.4
....
y_{24}	0	0	0

Từ ma trận đánh giá xác định theo (3), tác giả chuyển đổi biểu diễn mối quan hệ giữa các người dùng giả lập và sản phẩm dựa trên biểu diễn đồ thị $G = \langle V, E \rangle$. Trong đó, tập đỉnh V của đồ thị được chia thành ba tập: tập người dùng giả lập, tập sản phẩm ($V = Y \cup P$). Tập cạnh E của đồ thị được xác định theo công thức (4). Mỗi cạnh $e \in E$ đều có dạng $e = (i, t)$ trong đó $i \in Y, x \in P$. Trọng số của mỗi cạnh được xác định theo (5).

$$E = \{e = (i, x) \mid r_{ix} \neq \phi\} \quad (4)$$

$$w_{ix} = \begin{cases} r_{ix} & \text{if } (i, x) \in E \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Khi đó, đồ thị hai phía biểu diễn cho lọc cộng tác được thể hiện trong Hình 2.



Hình 2. Đồ thị biểu diễn cho lọc cộng tác gồm tập người dùng giả lập và tập sản phẩm

C. Dự đoán đánh giá của người dùng với sản phẩm cho lọc cộng tác dựa trên biểu diễn đồ thị

Dựa vào biểu diễn đồ thị đưa ra ở trên, bài toán dự đoán đánh giá của người dùng với sản phẩm cho lọc cộng tác được quy về bài toán tìm kiếm đường đi từ đỉnh người dùng y_i tới đỉnh sản phẩm p_x trên mô hình đồ thị cho trước. Do đồ thị biểu diễn đề xuất là đồ thị 2 phía, trong đó 1 phía là tập đỉnh người dùng, 1 phía là tập đỉnh sản phẩm nên với cách tiếp cận này, đánh giá dự đoán của người dùng y_i với sản phẩm p_x trên đồ thị đề xuất dựa trên việc tính toán tổng trọng số các đường đi có độ dài lẻ $L (L = 1, 3, 5 \dots)$ từ đỉnh y_i đến đỉnh p_x . Trong đó trọng số mỗi đường đi sẽ bằng tích trọng số của các cạnh tương ứng trên đường đi đó. Như vậy, đánh giá dự đoán của người dùng giả lập với sản phẩm sẽ được xác định dựa trên việc khai thác tất cả các mối quan hệ trực tiếp và gián tiếp giữa người dùng giả lập với sản phẩm đó trên đồ thị G . Đề xuất này được cải tiến từ nghiên cứu của Huang [22] cho tập dữ liệu có nhiều mức đánh giá và mở rộng ra cho bài toán lọc cộng tác theo ngữ cảnh.

Ví dụ: Đánh giá dự đoán của y_1 với p_3 sẽ được tính toán dựa trên tổng trọng số các đường đi từ y_1 tới p_3 . Điều này thu được bằng cách quan sát tất cả các đường đi từ y_1 tới p_3 , có 2 đường đi là $y_1 - p_1 - y_2 - p_3$ và $y_1 - p_1 - y_2 - p_2 - y_4$

– p_3 . Trọng số mỗi đường đi sẽ được tính bằng tích trọng số của các cạnh tương ứng trên đường đi đó, cụ thể đường đi $y_1 - p_1 - y_2 - p_3$ (độ dài 3) có trọng số là $1.0 * 0.8 * 0.4 = 0.32$, đường đi $y_1 - p_1 - y_2 - p_2 - y_4 - p_3$ có trọng số là $1.0 * 0.8 * 0.2 * 0.8 * 0.6 = 0.0768$, nên đánh giá dự đoán của y_1 với p_3 là $0.32 + 0.0768 = 0.3968$. Như vậy có thể thấy nếu ta mở rộng độ dài đường đi từ đỉnh người dùng tới đỉnh sản phẩm sẽ khai thác được khá nhiều mối quan hệ bậc cao giữa người dùng với sản phẩm, từ đó góp phần giải quyết hiệu quả vấn đề dữ liệu thưa của lọc cộng tác.

Tổng quát hóa, với dữ liệu đầu vào là ma trận $W(H \times M)$ được xác định theo công thức (5) thể hiện mối quan hệ giữa người dùng giả lập với sản phẩm, khi đó đánh giá dự đoán được xác định theo công thức (6). Trong đó $W^T(H, M)$ là ma trận chuyển vị của $W(H, M)$.

$$W^L(H \times M) = \begin{cases} W(H \times M) & \text{if } L=1 \\ W(H \times M) * W^T(H \times M) * W^{L-2}(H \times M) & \text{if } L=3,5,7,\dots \end{cases} \quad (6)$$

Ma trận $W(H \times M)$ cho biết biểu diễn mức độ phù hợp giữa tập người dùng giả lập và sản phẩm của hệ thống. Mức độ phù hợp này phụ thuộc vào độ dài đường đi L từ đỉnh người dùng giả lập đến đỉnh sản phẩm trên đồ thị. Các đường đi có độ dài lớn sẽ được đánh trọng số thấp, đường đi có độ dài nhỏ sẽ được đánh trọng số cao. Do vậy, ta cần xác định giá trị L để thực hiện tính toán bằng cách chọn giá trị L nhỏ nhất để $w_{ix}^L \neq 0$ với mọi $y_i \in Y$ và $p_x \in P$.

D. Sinh tư vấn

Từ mức độ phù hợp giữa người dùng giả lập và sản phẩm xác định theo (6), với mỗi người dùng cần tư vấn u_a tác giả chọn k sản phẩm có mức độ phù hợp cao nhất để tư vấn cho người dùng u_a .

Trên cơ sở các bước trình bày diễn giải trong các mục A, B, C, D cho bộ khung đề xuất triển khai phương pháp lọc trước theo ngữ cảnh dựa trên mô hình đồ thị cho hệ tư vấn cộng tác theo ngữ cảnh ở Hình 1, tác giả đề xuất thuật toán lọc cộng tác theo ngữ cảnh (CACF_USGraph) được miêu tả chi tiết trong Hình 3.

Đầu vào

- Ma trận đánh giá đa chiều (chứa thông tin ngữ cảnh)
- $u_a \in U$ là người dùng hiện thời cần được tư vấn
- $c_t \in (C_1 \times C_2 \times \dots \times C_K)$ là tình huống ngữ cảnh ứng với u_a
- L_{max} là giới hạn độ dài đường đi từ đỉnh người dùng tới đỉnh sản phẩm
- K là số lượng sản phẩm cần tư vấn cho u_a

Đầu ra:

- Danh sách K sản phẩm tư vấn tới người dùng u_a trong tình huống ngữ cảnh c_t .

Các bước tiến hành:

Bước 1. Chuyển đổi ma trận đánh giá dạng đa chiều R_1 về dạng hai chiều R_0
Theo thủ tục phân tách người dùng theo ngữ cảnh (Mục A).

Bước 2. Khởi tạo các ma trận $W(H \times M)$ theo công thức (5)

Bước 3. Dự đoán đánh giá của người dùng với sản phẩm

$L \leftarrow 2$; // Thiết lập độ dài đường đi ban đầu

Repeat

$$W^L(H \times M) = \begin{cases} W(H \times M) & \text{if } L=1 \\ W(H \times M) * W^T(H \times M) * W^{L-2}(H \times M) & \text{if } L=3,5,7,\dots \end{cases}$$

$L \leftarrow L + 2$; // Tăng độ dài đường đi.

Until ($w_{ix}^L \neq 0$ với mọi $i \in Y, x \in P$ hoặc $L=L_{max}$);

Bước 4. Sinh tư vấn cho người dùng hiện thời u_a trong ngữ cảnh c_t

- Gọi y_a là người dùng hiện thời giả lập được tạo ra khi kết hợp u_a và c_t .
- Căn cứ vào $W^L(H \times M)$ tính toán được từ bước 3, chọn K sản phẩm có đánh giá dự đoán cao nhất mà chưa được đánh giá bởi y_a để tư vấn cho y_a
- Tư vấn K sản phẩm trên cho người dùng hiện thời u_a trong tình huống ngữ cảnh c_t .

Hình 3. Thuật toán CACF_USGraph

IV. THỰC NGHIỆM VÀ KẾT QUẢ

A. Dữ liệu thực nghiệm

Để thấy rõ hiệu quả của phương pháp đề xuất, tác giả thực hiện tiến hành thực nghiệm trên hai bộ dữ liệu DepaulMovie, InCarMusic [24].

- Bộ dữ liệu DepaulMovie chứa 5043 đánh giá từ 97 người dùng cho 79 phim trong các tình huống ngữ cảnh khác nhau. Bộ dữ liệu này có 3 chiều ngữ cảnh là *Time*, *Location*, *Companion*. Chiều ngữ cảnh *Time* có 2 điều kiện ngữ cảnh (“Weekend”, “Weekday”), chiều ngữ cảnh *Location* có 2 điều kiện ngữ cảnh (“Home”, “Cinema”), chiều ngữ cảnh *Companion* có 3 điều kiện ngữ cảnh (“Alone”, “Family”, “Partner”). Các mức đánh giá nằm trong dải từ 1 đến 5, mức độ thưa thớt của dữ liệu là 94,516%. Các mức đánh giá 1, 2, 3, 4, 5 được chuyển đổi thành 0.2, 0.4, 0.6, 0.8, 1.0.
- Bộ dữ liệu InCarMusic chứa 3938 đánh giá từ 1042 người dùng, 139 album trong các tình huống ngữ cảnh khác nhau. Bộ dữ liệu này có 8 chiều ngữ cảnh là *Driving style*, *Road type*, *Landscape*, *Sleepiness*, *Traffic conditions*, *Mood*, *Weather*, *Natural Phenomena*. Chiều ngữ cảnh *Driving style* có 2 điều kiện ngữ cảnh (“Relaxed driving”, “Sport driving”), chiều ngữ cảnh *Road type* có 3 điều kiện ngữ cảnh (“City”, “Highway”, “Serpentine”), chiều ngữ cảnh *Landscape* có 4 điều kiện ngữ cảnh (“Coast line”, “country side”, “mountains/hills”, “Urban”), chiều ngữ cảnh *Sleepiness* có 2 điều kiện ngữ cảnh (“Awake”, “Sleepy”), chiều ngữ cảnh *Traffic*

conditions có 3 điều kiện ngữ cảnh (“Free road”, “Many Cars”, “Traffic jam”), chiều ngữ cảnh *Mood* có 4 điều kiện ngữ cảnh (“Active”, “Happy”, “Lazy”, “Sad”), chiều ngữ cảnh *Weather* có 4 điều kiện ngữ cảnh (“Cloudy”, “Snowing”, “Sunny”, “Rainy”), chiều ngữ cảnh *Natural Phenomena* có 4 điều kiện ngữ cảnh (“Day time”, “Morning”, “Night”, “Afternoon”). Các mức đánh giá nằm trong dải từ 1 đến 5, mức độ thưa thớt của dữ liệu là 99.9996996%. Các mức đánh giá 1, 2, 3, 4, 5 được chuyển đổi thành 0.2, 0.4, 0.6, 0.8, 1.0.

B. Cài đặt thực nghiệm

1) Độ đo

Hai nhiệm vụ chính của hệ tư vấn là dự đoán đánh giá và tư vấn danh sách ngắn các sản phẩm cho người dùng hiện thời. Để đánh giá hiệu quả của đánh giá dự đoán, các độ đo thường được sử dụng là *MAE, RMSE, MPE* [1, 25]. Để đánh giá hiệu quả tư vấn danh sách sản phẩm, các độ đo điển hình được sử dụng là *Precision@N, Recall@N* và *F – measure@N*. Trong bài báo này, tác giả tập trung đánh giá hiệu quả tư vấn danh sách sản phẩm của phương pháp đề xuất trong sự so sánh với các phương pháp tư vấn theo ngữ cảnh cơ sở nên độ đo *Precision@N, Recall@N* và *F – measure@N* [25] sẽ được lựa chọn để đánh giá kết quả. Cụ thể độ đo như sau:

- Độ chính xác *Precision@N* cho biết tỷ lệ dự đoán chính xác trong top-N sản phẩm dự đoán cho mỗi người dùng (top-N items).

$$Precision@N = \frac{|{\text{relevant items}} \cap {\text{top-N items}}|}{N} \quad (7)$$

Độ nhạy *Recall@N* cho biết tỷ lệ dự đoán chính xác trong số những sản phẩm thực tế thích bởi người dùng (relevant items)

$$Recall@N = \frac{|{\text{relevant items}} \cap {\text{top - N items}}|}{|{\text{relevant items}}|} \quad (8)$$

- Để cân bằng giữa hai độ đo *Precision@N* và *Recall@N*, một độ đo được đưa ra đó là *F – measure@N* được xác định theo (9). Giá trị F-measure lớn thể hiện thuật toán tư vấn có độ chính xác càng cao.

$$F\text{-measure@N} = \frac{2}{\frac{1}{Precision@N} + \frac{1}{Recall@N}} \quad (9)$$

2) Phương pháp thực nghiệm

Để đánh giá độ chính xác của danh sách sản phẩm tư vấn, tác giả thực hiện phân chia tập dữ liệu *U* thành 2 tập *U_{train}* và *U_{test}* sử dụng phương pháp kiểm thử chéo (k-fold cross-validation) vì đây là phương pháp được sử dụng rộng rãi và cho kết quả đánh giá khách quan nhất. Ý tưởng của k-fold cross-validation là chia tập dữ liệu *U* thành *k* tập dữ liệu nhỏ (folds) có kích thước như nhau, sau đó chúng ta tiến hành kiểm nghiệm thuật toán *k* lần và lấy trung bình kết quả các lần kiểm nghiệm đó để ra kết quả kiểm nghiệm cuối cùng. Tại mỗi lần kiểm nghiệm sẽ lấy 1 trong *k* tập dữ liệu nhỏ làm tập *U_{test}*, các tập dữ liệu nhỏ còn lại được dùng làm tập *U_{train}*. Với mỗi người dùng $u_a \in U_{test}$, một số đánh

giá r_{u_a} được chọn ngẫu nhiên sẽ bị che đi, những giá trị không bị che đi sẽ kết hợp với tập *U_{train}* để thực hiện huấn luyện theo thuật toán tư vấn. Kết quả tư vấn được cho mỗi người dùng trong tập *U_{test}* sau quá trình huấn luyện sẽ được đối chiếu với các đánh giá ban đầu để xác định tính chính xác của thuật toán tư vấn. Trong thực nghiệm, tác giả sẽ lấy *k = 10* để tiến hành chia dữ liệu kiểm nghiệm. Việc thực nghiệm được thực hiện 10 lần và lấy trung bình kết quả thực nghiệm để đảm bảo tính khách quan của kết quả thực nghiệm.

3) Các phương pháp tư vấn được sử dụng để so sánh

Chúng tôi tiến hành kiểm nghiệm thuật toán đề xuất (CACF_GRAPH) trong sự so sánh với các phương pháp tư vấn cộng tác theo ngữ cảnh cơ sở sau:

- *CPTF*: Sử dụng mô hình Matrix Factorization thông qua kỹ thuật phân tích nhân tố tiềm ẩn (Tensor Factorization – TF)
- *CAMF_C* (Context-aware matrix factorization-Context): Sử dụng phương pháp mô hình hóa ngữ cảnh độc lập dựa trên sự chênh lệch đánh giá giữa các điều kiện ngữ cảnh.
- *CAMF_CU* (Context-aware matrix factorization - User base context): Sử dụng phương pháp mô hình hóa ngữ cảnh độc lập dựa trên sự chênh lệch đánh giá của người dùng giữa các điều kiện ngữ cảnh.
- *CAMF_CI* (Context-aware matrix factorization - Item base context): Sử dụng phương pháp mô hình hóa ngữ cảnh độc lập dựa trên sự chênh lệch đánh giá đối với sản phẩm giữa các điều kiện ngữ cảnh.
- *CSLIM_C*: Sử dụng phương pháp tuyến tính thưa SLIM kết hợp với mô hình hóa ngữ cảnh dựa vào độ lệch đánh giá.
- *CSLIM_IC* (Contextual Sparse Linear Method - Independent Context Similarity): Sử dụng phương pháp tuyến tính thưa SLIM kết hợp với mô hình hóa ngữ cảnh dựa trên độ tương quan ngữ cảnh độc lập.
- *CSLIM_MCS* (Contextual Sparse Linear Method - Multidimensional Context Similarity): Sử dụng phương pháp tuyến tính thưa SLIM kết hợp với mô hình hóa ngữ cảnh dựa trên độ tương quan đa chiều.
- *UserSplitting-BiasedMF*: Sử dụng phương pháp phân chia sản phẩm theo ngữ cảnh, sau đó huấn luyện và đưa ra tư vấn sử dụng phương pháp phân rã ma trận BiasedMF.
- *ItemSplitting-BiasedMF*: Sử dụng phương pháp phân chia người dùng theo ngữ cảnh, sau đó huấn luyện và đưa ra tư vấn sử dụng phương pháp phân rã ma trận BiasedMF.
- *UISplitting-BiasedMF*: Sử dụng phương pháp phân chia cả người dùng và sản phẩm theo ngữ cảnh, sau đó huấn luyện và đưa ra tư vấn sử dụng phương pháp phân rã ma trận BiasedMF.

C. Kết quả thực nghiệm

Bảng 4. Kết quả kiểm nghiệm của các phương pháp lọc cộng tác theo ngữ cảnh cho bộ dữ liệu DepaulMovie

Phương pháp	Precision @10	Recall @10	F-measure@10
CPTF	0.051	0.181	0.080
CAMF_C	0.088	0.323	0.138
CAMF_CU	0.078	0.280	0.122
CAMF_CI	0.067	0.236	0.104
CSLIM_C	0.063	0.214	0.097
CSLIM_ICS	0.065	0.218	0.100
CSLIM_MCS	0.085	0.315	0.134
UserSplitting-BiasedMF	0.089	0.324	0.140
ItemSplitting-BiasedMF	0.085	0.316	0.134
UISplitting-BiasedMF	0.081	0.298	0.127
CACF_USGraph	0.103	0.327	0.157

Bảng 5. Kết quả kiểm nghiệm của các phương pháp lọc cộng tác theo ngữ cảnh cho bộ dữ liệu InCarMusic

Phương pháp	Precision @10	Recall @10	F-measure@10
CPTF	0.017	0.104	0.0292
CAMF_C	0.032	0.218	0.0558
CAMF_CU	0.027	0.177	0.0469
CAMF_CI	0.015	0.087	0.0256
CSLIM_C	0.018	0.127	0.0315
CSLIM_ICS	0.039	0.231	0.0667
CSLIM_MCS	0.018	0.106	0.0308
UserSplitting-BiasedMF	0.031	0.215	0.0542
ItemSplitting-BiasedMF	0.033	0.226	0.0576
UISplitting-BiasedMF	0.032	0.219	0.0558
CACF_USGraph	0.043	0.242	0.073

Một số nhận xét được đưa ra căn cứ vào phân tích kết quả thực nghiệm đưa ra trong Bảng 4, Bảng 5 như sau:

- Các phương pháp lọc trước theo ngữ cảnh trên cơ sở sử dụng phương pháp phân tách sản phẩm/ người dùng nguyên thủy (Sử dụng một chiều ngữ cảnh) và phương pháp phân rã ma trận (ItemSplitting-BiasedMF, UserSplitting-BiasedMF, UISplitting-BiasedMF) cho lại độ chính xác tốt hơn các phương pháp mô hình hóa ngữ cảnh dựa trên độ tương quan ngữ cảnh độc lập. Điều này chứng tỏ tiếp cận lọc trước ngữ cảnh theo hướng phân tách sản phẩm/ người dùng cho phép tích hợp thông tin ngữ cảnh giúp nâng cao hiệu quả tư vấn.
- Phương pháp đề xuất CACF_USGraph cho lại độ chính xác tốt hơn so với việc kết hợp các phương pháp lọc trước theo ngữ cảnh theo hướng sử dụng phương pháp phân tách sản phẩm/ người dùng nguyên thủy và phương pháp phân rã ma trận. Điều này có thể lý giải là do thủ tục phân tách theo người dùng cải tiến cho phép tích hợp đầy đủ thông tin ngữ cảnh và vấn đề dữ

liệu thừa được giải quyết bằng việc khai thác đầy đủ mối quan hệ trực tiếp và gián tiếp giữa người dùng và sản phẩm vào quá trình huấn luyện và dự đoán đánh giá, từ đó nâng cao hiệu quả tư vấn, đây chính là lợi thế của phương pháp đề xuất so với các phương pháp cơ sở.

V. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Bài báo đã trình bày đề xuất một phương pháp lọc cộng tác theo ngữ cảnh mới cho hệ tư vấn. Phương pháp đề xuất cho phép tích hợp đầy đủ thông tin ngữ cảnh và giải quyết hiệu quả vấn đề dữ liệu thừa. Trong đó, việc tích hợp ngữ cảnh được thực hiện bằng thủ tục phân tách người dùng theo ngữ cảnh cải tiến nhằm khắc phục hạn chế của phương pháp phân tách người dùng theo ngữ cảnh nguyên thủy, đó là cho phép tích hợp đầy đủ thông tin ngữ cảnh trong việc chuyển hóa người dùng ban đầu thành người dùng giả lập. Quá trình phân tách người dùng theo ngữ cảnh sẽ biến đổi ma trận đánh giá đa chiều vốn dĩ đã thừa về ma trận đánh giá hai chiều càng trở lên thừa thớt hơn nữa. Để giải quyết vấn đề dữ liệu thừa này, tác giả tiếp cận phương pháp biểu diễn và xây dựng thuật toán dự đoán dựa trên mô hình đồ thị nhằm khai thác đầy đủ mối quan hệ trực tiếp và gián tiếp giữa các người dùng và sản phẩm. Kết quả thực nghiệm trên một số bộ dữ liệu thực cho thấy phương pháp đề xuất cải thiện đáng kể chất lượng dự đoán so với các phương pháp lọc cộng tác theo ngữ cảnh cơ sở trước đây.

Trong thời gian tới, tác giả dự định sẽ mở rộng nghiên cứu của mình cho hệ tư vấn theo ngữ cảnh nhằm khai thác được sự tương quan giữa các tình huống ngữ cảnh phục vụ cho quá trình huấn luyện nhằm nâng cao chất lượng tư vấn.

TÀI LIỆU THAM KHẢO

- [1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005, doi: 10.1109/TKDE.2005.99.
- [2] G. Adomavicius, B. Mobasher, F. Ricci, and A. Tuzhilin, "Context-Aware Recommender Systems," *AI Mag.*, vol. 32, no. 3, pp. 67–80, 2011.
- [3] L. Baltrunas, B. Ludwig, and F. Ricci, "Matrix Factorization Techniques for Context Aware," *Acm Rs*, no. October, pp. 301–304, 2011, doi: 10.1145/2043932.2043988.
- [4] L. Cai, J. Xu, J. Liu, and T. Pei, "Integrating spatial and temporal contexts into a factorization model for POI recommendation," *Int. J. Geogr. Inf. Sci.*, vol. 32, no. 3, pp. 524–546, 2018, doi: 10.1080/13658816.2017.1400550.
- [5] A. Razia Sulthana and S. Ramasamy, "Ontology and context based recommendation system using Neuro-Fuzzy Classification," *Comput. Electr. Eng.*, vol. 0, pp. 1–13, 2018, doi: 10.1016/j.compeleceng.2018.01.034.
- [6] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, *Recommender systems handbook*. Springer, 2011.
- [7] Haruna, Khalid & Ismail, Maizatul Akmar & Suhendroyono, Suhendroyono & Damiasih, Damiasih & Pierewan, Adi & Chiroma, Haruna & Herawan, Tutut. (2017). Context-Aware Recommender System: A Review of Recent Developmental Process and Future Research Direction. *Applied Sciences*. 7. 1211. 10.3390/app7121211.
- [8] L. Baltrunas and F. Ricci, "Context-Based Splitting of Item Ratings in Collaborative Filtering," in *Proceedings of the third ACM conference on Recommender systems - RecSys*

- '09, 2009, pp. 245–248.
- [9] L. Baltrunas, B. Ludwig, F. Ricci. Matrix factorization techniques for context aware recommendation. *ACM RecSys*, 2011
- [10] Y. Zheng, B. Mobasher, R. Burke. CSLIM: Contextual SLIM Recommendation Algorithms. *ACM RecSys*, 2014
- [11] Y. Zheng, R. Burke, and B. Mobasher, “Splitting approaches for context-aware recommendation,” *Proc. 29th Annu. ACM Symp. Appl. Comput. - SAC '14*, pp. 274–279, 2014, doi: 10.1145/2554850.2554989.
- [12] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, “Multiverse Recommendation: N-dimensional Tensor Factorization for Context-aware Collaborative Filtering,” in *Proceedings of the Fourth ACM Conference on Recommender Systems*, 2010, pp. 79–86, doi: 10.1145/1864708.1864727.
- [13] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin, “Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach,” *ACM Trans. Inf. Syst.*, vol. 23, no. 1, pp. 103–145, Jan. 2005, doi: 10.1145/1055709.1055714.
- [14] Y. Zheng, “Tutorial: Context In Recommender Systems,” 2016.
- [15] Y. Zheng, B. Mobasher, R. Burke. Deviation-Based Contextual SLIM Recommenders. *ACM CIKM*, 2014
- [16] S. Lee, S.-I. Song, M. Kahng, D. Lee, and S.-G. Lee. Random walk based entity ranking on graph for multi-dimensional recommendation. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*, pp. 93–100, ACM, October 2011.
- [17] Şamdan, Emrah & Taşcı, Arda & Cicekli, Nihan. (2014). A Graph-based Collaborative and Context-aware Recommendation system for TV programs. *RecSys 2014 TV Workshop*.
- [18] Z. Bahramian, R. Ali Abbaspour, and C. Claramunt. A Context-Aware Tourism Recommender System Based On A Spreading Activation Method. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLII-4/W4, 2017 Tehran's Joint ISPRS Conferences of GI Research, SMPR and EOEC 2017, 7–10 October 2017, Tehran, Iran.
- [19] Neves ARM, Carvalho AMG, Ralha CG (2013). Agent-Based Architecture For Context-Aware And Personalized Event Recommendation. *Expert Systems with Applications* 41(2):563–573
- [20] Bedi P, Singh R (2015). User Interest Expansion Using Spreading Activation For Generating Recommendations. In: *2015 International conference on advances in computing, communications and informatics*, Kerala, pp 766–771
- [21] Yong Zheng. Tutorial: Context In Recommender Systems. Conference: Proceedings of the 31st ACM SIGAPP Symposium on Applied Computing (ACM SAC 2016), Pisa, Italy, April 2016
- [22] Z. Huang, D. Zeng, H. Chen, “Analyzing Consumer-product Graphs: Empirical Findings and Applications in Recommender Systems”, *Management Science*, 53(7), 1146–1164 (2007).
- [23] C.C. Aggarwal, J.L. Wolf, K.L. Wu, and P.S. Yu, “Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering”, *Proc. Fifth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining* (1999).
- [24] Y. Zheng, B. Mobasher, and R. Burke, “CARSKit: A Java-Based Context-Aware Recommendation Engine,” in *Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, 2015, pp. 1668–1671, doi: 10.1109/ICDMW.2015.222.
- [25] J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J.T. Riedl. Evaluating Collaborative Filtering Recommender Systems. *ACM Trans. Information Systems*, vol. 22, No. 1, pp. 5–53, 2004.
- [26] L. Baltrunas and F. Ricci. Experimental evaluation of context-dependent collaborative filtering using item splitting. *User Modeling and User-Adapted Interaction*. Vol.24, no. 1-2, pp. 7–34, 2014.
- [27] Y. Zheng, R. Burke, and B. Mobasher. Differential context relaxation for context-aware travel recommendation. In *Proceedings of the 13th International Conference on Electronic Commerce and Web Technologies (EC-WEB '12)*, pp. 88–99, 2012
- [28] V. Codina, F. Ricci, and L. Ceccaroni. Exploiting the semantic similarity of contextual situations for pre-filtering recommendation. In *Proceedings of the 21th International Conference on User Modeling, Adaptation, and Personalization*, pp. 165–177, Springer, 2013.

A CONTEXTUAL MODELING METHOD FOR CONTEXT-AWARE RECOMMENDER SYSTEMS

Abstract: Collaborative filtering is the typical approach in building recommender systems. Basically, the traditional collaborative filtering methods mainly focus on exploiting only a rating matrix between users and items to perform training and recommend new suitable items to an active user. However, user's ratings for items often change according to different contexts, so the context-aware collaborative filtering has been proven to improve quality of suggestions significantly in many practical applications compared to the traditional collaborative filtering approaches. There are many different proposed methods to solve the context-aware collaborative filtering, however this field still has many challenges, in which the main difficulty of context-aware recommender systems is how to integrate effectively and the data sparseness problem which directly affect to quality of the recommendation. In this paper, I will propose a new context-aware collaborative filtering that allow fully integrated context situations and resolve effectively the data sparseness problem. The proposed method is based on contextual pre-filtering approach. In there, the problem of context integration is done by a advanced user splitting technique and the sparse data problem is solved by fully exploiting the direct and indirect relationships between users and items during training process based on the graph model. The experimental results on some real data sets show that the proposed method outperforms baseline contextual modeling methods for context-aware recommendation systems.

Keywords: Context-aware collaborative filtering - CACF; Context; Context-aware recommender system - CARS; Contextual pre-filtering; User splitting

Đỗ Thị Liên, Nhận bằng tốt nghiệp đại học, thạc sĩ và học vị tiến sĩ tại Học viện Công nghệ Bưu chính Viễn thông vào các năm 2010, 2013, 2020. Hiện là giảng viên tại Học Viện Công nghệ Bưu Chính Viễn Thông.

Lĩnh vực nghiên cứu chính: học máy ứng dụng trong lọc thông tin, khai phá dữ liệu, phát triển ứng dụng đa phương tiện.

