

MỘT PHƯƠNG PHÁP TƯ VẤN CỘNG TÁC THEO NGỮ CẢNH

Đỗ Thị Liên, Nguyễn Duy Phương

¹Học viện Công nghệ Bưu chính Viễn thông
liendt@ptit.edu.vn, phuongnd@ptit.edu.vn

TÓM TẮT: Lọc cộng tác (collaborative filtering) là phương pháp phổ biến được dùng trong xây dựng các hệ tư vấn. Các phương pháp lọc cộng tác hiện nay khai thác duy nhất tập dữ liệu đánh giá người dùng đối với sản phẩm để tìm ra cộng đồng người dùng có cùng chung sở thích và tiến hành dự đoán những sản phẩm phù hợp cho mỗi người dùng. Trong bài báo này, chúng tôi đề xuất một phương pháp lọc cộng tác theo ngữ cảnh sử dụng sản phẩm của mỗi người dùng. Phương pháp được tiến hành bằng cách biểu diễn mối quan hệ giữa các người dùng, sản phẩm và ngữ cảnh trên cùng một mô hình đồ thị. Trên cơ sở biểu diễn đồ thị cho hệ tư vấn cộng tác theo ngữ cảnh, chúng tôi xây dựng mô hình dự đoán mức độ phù hợp của sản phẩm với người dùng trong từng tình huống ngữ cảnh cụ thể. Kết quả thử 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 trước đây.

Từ khóa (Keywords): Hệ tư vấn dựa vào ngữ cảnh (Context-aware recommender system - CARS); Lọc cộng tác dựa vào ngữ cảnh (Context-aware collaborative filtering - CACF); Ngữ cảnh (Context); Mô hình hóa ngữ cảnh dựa trên đồ thị (Graph-based contextual modeling).

I. MỞ ĐẦU

Hệ tư vấn (Recommender System) được xem như một hệ thống lọc tích cực, có chức năng hỗ trợ đưa ra quyết định, nhằm mục đích cung cấp cho người sử dụng những gợi ý về thông tin, sản phẩm và dịch vụ phù hợp nhất với yêu cầu và sở thích riêng của từng người. Hệ tư vấn truyền thống chỉ quan tâm tới hai đối tượng của hệ tư vấn là người dùng (user) và sản phẩm (item). Tuy nhiên trên thực tế, sở thích của người dùng lại không cố định. Ví dụ một người trời nóng thì thích ăn kem, uống sinh tố, nhưng khi trời lạnh lại thích ăn phở, uống cà phê nóng. Hoặc cùng một bộ phim nhưng trời mưa thì thích xem còn trời khô ráo thì có khi lại không thích. Có thể nói sở thích của người dùng bị tác động nhiều bởi những yếu tố ngữ cảnh bên ngoài. 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 một số hệ tư vấn dựa trên ngữ cảnh như hệ tư vấn địa điểm du lịch [3], thức ăn [4], phim ảnh [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”. Chính vì vậy, yếu tố ngữ cảnh được nghiên cứu rộng rãi trong lĩnh vực hệ tư vấn. 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). Những yếu tố này hoàn toàn có thể ảnh hưởng tới quyết định chọn địa điểm du lịch của người dùng. Hệ tư vấn sẽ đóng vai trò ghi nhớ lại sở thích của người dùng theo ngữ cảnh để đưa ra những gợi ý chính xác nhất.

Mặc dù đã có một số đề xuất được đưa ra để giải quyết bài toán tư vấn dựa vào ngữ cảnh, nhưng làm thế nào để tích hợp hiệu quả thông tin ngữ cảnh vào hệ tư vấn vẫn là vấn đề nghiên cứu mở, có tính thời sự và thu hút được nhiều quan tâm của cộng đồng nghiên cứu. Các kết quả nghiên cứu cũng chỉ ra rằng việc lựa chọn mở rộng các phương pháp tư vấn truyền thống phù hợp sẽ ảnh hưởng đáng kể tới chất lượng của hệ tư vấn dựa vào ngữ cảnh [2]. Trong bài báo này, chúng tôi đề xuất một phương pháp tư vấn cộng tác theo ngữ cảnh mới, lấy phương pháp lọc cộng tác là cơ sở để thực hiện quá trình huấn luyện dữ liệu. Phương pháp được tiến hành bằng cách biểu diễn tất cả mối quan hệ giữa các người dùng, sản phẩm và ngữ cảnh trên cùng một mô hình đồ thị hợp nhất. Khi đó việc xem xét mức độ phù hợp của sản phẩm với người dùng trong từng tình huống ngữ cảnh cụ thể dựa trên quá trình tìm kiếm tất cả các đường đi từ đỉnh người dùng tới đỉnh sản phẩm đi qua đỉnh tình huống ngữ cảnh. Trên cơ sở biểu diễn đồ thị cho hệ tư vấn cộng tác theo ngữ cảnh, chúng tôi xây dựng mô hình dự đoán mức độ phù hợp của sản phẩm với người dùng trong từng điều kiện ngữ cảnh cụ thể. Kết quả thử 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 trước đây. Đề trọng tâm vào phương pháp đề xuất, phần II chúng tôi trình bày bài toán lọc cộng tác theo ngữ cảnh. Tiếp đến là phương pháp đề xuất về lọc cộng tác theo ngữ cảnh trên đồ thị trong phần III. Phần IV trình bày phương pháp thử nghiệm và đánh giá. Phần V nêu kết luận và hướng phát triển trong thời gian tới.

II. BÀI TOÁN TƯ VẤN CỘNG TÁC THEO NGỮ CẢNH (CONTEXT-AWARE COLLABORATIVE FILTERING - CACF)

Đối với hệ tư vấn cộng tác truyền thống, ta chỉ quan tâm tới mối quan hệ giữa hai nhóm đối tượng là người dùng và sản phẩm để thực hiện huấn luyện và đưa ra dự đoán. Ta có thể mô phỏng bài toán tư vấn cộng tác truyền thống dựa trên ma trận đánh giá hai chiều sau:

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

Trong khi đó, đối với hệ tư vấn cộng tác tích hợp yếu tố ngữ cảnh, ngoài thông tin về hai đối tượng người dùng và sản phẩm, hệ thống còn quan tâm tới những yếu tố ngữ cảnh khi người dùng đánh giá một sản phẩm để phục vụ cho quá trình huấn luyện và tư vấn. Khi đó bài toán tư vấn cộng tác tích hợp yếu tố ngữ cảnh lúc này sẽ có dạng ma trận đánh giá đa chiều như sau:

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

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 hệ tư vấn cộng tác theo ngữ cảnh là dự đoán đánh giá và đưa ra tư vấn các sản phẩm mới cho người dùng trong 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 3 người dùng $U = \{u_1, u_2, u_3\}$, 2 sản phẩm $P = \{p_1, p_2\}$, 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

User	Item	Rating	Time	Location	Companion
u_1	p_1	5	Weekend	Home	Kids
u_1	p_1	4	Weekday	Home	Family
u_2	p_1	3	Weekend	Cinema	Partner
u_2	p_1	4	Weekday	Home	Family
u_3	p_1	3	Weekend	Cinema	Partner
u_3	p_2	2	Weekend	Cinema	Partner

Một số thuật ngữ quy ước được sử dụng trong các hệ tư vấn dựa vào ngữ cảnh được biết đến đó 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 một giá trị của một chiều ngữ cảnh được đánh giá bởi người dùng, ví dụ: chiều ngữ cảnh “Time” có thể nhận 1 trong 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 (“Kids”, “Family”, “Partner”). Thuật ngữ *tình huống ngữ cảnh* chỉ một bộ giá trị điều kiện ngữ cảnh của các chiều ngữ cảnh tương ứng được đánh giá bởi người dùng, ví dụ người dùng u_1 đánh giá sản phẩm p_1 trong tình huống ngữ cảnh (“Weekend”, “Home”, “Kids”) là 5.

Có một số phương pháp đề xuất khác nhau để giải quyết bài toán tư vấn cộng tác theo ngữ cảnh. Các phương pháp này thuộc ba hướng: (i) lọc trước ngữ cảnh, (ii) lọc sau ngữ cảnh và (iii) mô hình hóa dựa vào ngữ cảnh.

Về cơ bản, các phương pháp tư vấn cộng tác theo ngữ cảnh thuộc hướng lọc trước ngữ cảnh (i) 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. Một số phương pháp lọc tập dữ liệu ban đầu được đưa ra như: Phương pháp phân chia dữ liệu (Splitting) [9], phương pháp sử dụng một tập con các chiều ngữ cảnh để lọc (Context Relaxation) [10], phương pháp lọc dựa trên ngữ nghĩa (Semantic Filtering) [11]. 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, ... [2, 14, 15, 24] 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.

Hướng thứ hai để giải quyết bài toán tư vấn cộng tác theo ngữ cảnh là lọc sau ngữ cảnh (ii). 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 tư vấn sẽ được lọc lại một lần nữa 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 (i).

Hướng thứ ba để giải quyết bài toán tư vấn cộng tác theo ngữ cảnh là mô hình hóa dựa vào ngữ cảnh (iii). Theo hướng này 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 và mô hình hóa ngữ cảnh phụ thuộc [5]. Tensor Decomposition [12] là một phương pháp điển hình thuộc nhóm phương pháp mô hình hóa ngữ cảnh độc lập. Tensor Decomposition 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 [13]. 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. Hai kỹ thuật được sử dụng trong phương pháp này đó là mô hình hóa dựa trên độ chênh lệch và mô hình hóa dựa trên độ tương tự [22]. Kỹ thuật mô hình hóa dựa trên độ chênh lệch sẽ thiết lập một mức chênh lệch cộng thêm vào đánh giá không có ngữ cảnh để suy ra đánh giá của người dùng cho sản phẩm trong một tình huống ngữ cảnh cụ thể. Kỹ thuật mô hình hóa dựa trên độ tương tự sẽ thiết lập một mức độ tương tự nhân với đánh giá không có ngữ cảnh để

điều chỉnh đánh giá của người dùng cho sản phẩm trong một tình huống ngữ cảnh cụ thể. Theo đó, để đưa ra dự đoán đánh giá của người dùng với sản phẩm trong từng tình huống ngữ cảnh, có hai việc cần thực hiện đó là: (1) Lựa chọn phương pháp lọc cộng tác để dự đoán đánh giá của người dùng với sản phẩm khi không có ngữ cảnh và (2) Lựa chọn độ đo tính mức chênh lệch hoặc mức tương tự giữa các tình huống ngữ cảnh. Với việc (1) về cơ bản chúng ta có thể sử dụng tất cả các phương pháp lọc cộng tác truyền thống để thực hiện, hai phương pháp tư vấn cộng tác theo ngữ cảnh dựa trên Matrix Factorization [14] và SLIM [15, 16] được đánh giá là mang lại hiệu quả tương đối tốt. Với việc (2) chúng ta có thể sử dụng những độ đo khoảng cách (Euclid, Minkowski,...), độ đo tương tự (Cosin, Entropy,...), độ đo tương quan (Pearson, Root Mean Square, Spearman Rank, Kendal,...) để tính toán mức độ chênh lệch hoặc tương tự giữa các tình huống ngữ cảnh. 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 [8]. 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 dựa trên các giải thuật tư vấn truyền thống như Matrix Factorization, SLIM,... là vấn đề dữ liệu thưa và khả năng mở rộng của nó.

Một hướng tiếp cận khác để tích hợp ngữ cảnh vào hệ tư vấn cộng tác là dựa trên mô hình đồ thị. Nhiều nghiên cứu đã chỉ ra rằng mô hình đồ thị giải quyết khá tốt vấn đề dữ liệu thưa và khả năng mở rộng dữ liệu [17]. Neves ARM và các cộng sự của mình [20] đã đưa ra đề xuất phương pháp tư vấn theo ngữ cảnh trên cơ sở Ontology và kỹ thuật kích hoạt lan truyền (Spreading Activation). Obitko [21] đề xuất một phương pháp tiếp cận mới nâng cao hiệu quả tư vấn cho người dùng trong các nhà hàng dựa vào Ontology và kỹ thuật kích hoạt lan truyền. Emrah và các cộng sự [18] đề xuất một hệ tư vấn theo ngữ cảnh dựa trên một mô hình đồ thị cộng tác cho các chương trình trên ti vi. Z. Bahramian và các cộng sự [19] đề xuất hệ tư vấn các địa điểm du lịch theo ngữ cảnh dựa trên một mô hình kích hoạt lan truyền. Tuy nhiên các nghiên cứu đã có này chủ yếu tập trung vào việc xây dựng các mô hình biểu diễn đồ thị áp dụng riêng lẻ cho từng bài toán tư vấn theo ngữ cảnh cụ thể, mà chưa có một giải pháp tổng thể chung cho các bài toán tư vấn theo ngữ cảnh.

Để giảm thiểu những hạn chế nêu trên, chúng tôi đề xuất một phương pháp tư vấn cộng tác theo ngữ cảnh mới dựa trên mô hình đồ thị hợp nhất, để thực hiện huấn luyện và dự đoán mức độ phù hợp của sản phẩm với người dùng trong từng tình huống ngữ cảnh cụ thể. Nội dung cụ thể của phương pháp đề xuất được trình bày chi tiết trong Mục 3.

III. ĐỀ XUẤT PHƯƠNG PHÁP TÍNH MỨC ĐỘ PHÙ HỢP CỦA SẢN PHẨM VỚI NGƯỜI DÙNG TRONG TÌNH HUỐNG NGỮ CẢNH CỤ THỂ DỰA TRÊN MÔ HÌNH ĐỒ THỊ

Có nhiều đề xuất khác nhau giải quyết bài toán lọc cộng tác trên mô hình đồ thị cho hệ tư vấn truyền thống nói chung [1, 23, 25] và một số đề xuất giải quyết bài toán lọc cộng tác theo ngữ cảnh nói riêng như đề cập trong Mục 2. Về bản chất các phương pháp đều tiến hành giải quyết ba vấn đề: *Phương pháp biểu diễn đồ thị cho bài toán tư vấn cộng tác, phương pháp tính toán mức độ phù hợp của người dùng với sản phẩm và xây dựng thuật toán dự đoán quan điểm của người dùng đối với các sản phẩm dựa vào đồ thị*. Trong bài báo này, chúng tôi giải quyết ba vấn đề này cho hệ tư vấn cộng tác theo ngữ cảnh dựa vào mô hình đồ thị bằng cách mở rộng mô hình biểu diễn và thuật toán dự đoán trên đồ thị đã được Huang đề xuất trong [23] có tích hợp yếu tố ngữ cảnh trên đó. Tiếp đến chúng tôi đề xuất phương pháp xác định mức độ phù hợp của sản phẩm với người dùng trong tình huống ngữ cảnh cụ thể theo các kỹ thuật kích hoạt lan truyền trên đồ thị hợp nhất. Đây là bước cơ sở quan trọng quyết định hiệu quả của các phương pháp tư vấn cộng tác theo ngữ cảnh. Phương pháp đề xuất đã cho lại kết quả dự đoán tốt hơn so với các phương pháp tư vấn cộng tác theo ngữ cảnh cơ sở.

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

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. Khi đó mỗi tình huống ngữ cảnh được ký hiệu là $c_t \in C$ với chiều ngữ cảnh C được định nghĩa để đại diện cho K chiều ngữ cảnh, $C = C_1 \times C_2 \times \dots \times C_K$.

Ma trận đánh giá đa chiều $R=(r_{itx})$ thể hiện đánh giá của người dùng $u_i \in U$ với sản phẩm $p_x \in P$ trong từng tình huống ngữ cảnh cụ thể $c_t \in C$ là đầu vào của hệ tư vấn cộng tác theo ngữ cảnh. Trong đó, $i=1 \dots N$, $x=1 \dots M$, t nhận giá trị trong dải từ 1 tới số lượng tình huống ngữ cảnh trong C . Không hạn chế tính tổng quát của bài toán, ta giả sử $r_{itx} = v$ nếu người dùng $u_i \in U$ đánh giá sản phẩm $p_x \in P$ trong tình huống ngữ cảnh $c_t \in C$ với mức độ v , trong đó $v \in [0, 1]$.

Để thuận tiện trong trình bày, ta viết $u_i \in U$ ngắn gọn thành $i \in U$, $p_x \in P$ ngắn gọn thành $x \in P$ và $c_t \in C$ là $t \in C$.

$$r_{itx} = \begin{cases} v, & \text{Nếu người dùng } i \text{ thích sản phẩm } x \text{ trong tình huống ngữ cảnh } t \text{ ở mức độ } v \text{ (} 0 \leq v \leq 1 \text{)}. \\ \emptyset, & \text{Nếu người dùng } i \text{ chưa biết hoặc chưa đánh giá sản phẩm } x. \end{cases} \quad (3)$$

Biểu diễn ma trận đánh giá theo (3) sẽ không ảnh hưởng đến các hệ thống tư vấn cộng tác theo ngữ cảnh 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 $r_{itx} \in \{1, 2, \dots, V\}$, ta chỉ cần thực hiện phép biến đổi đơn giản chuyển $r_{itx} = \frac{r_{itx}}{V}$. Phép biến đổi này vẫn bảo toàn được

mức độ 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 [23]. Ví dụ với hệ lọc cộng tác được cho trong Bảng 1, sẽ được chuyển đổi biểu diễn theo (3) thành Bảng 2, trong đó các tình huống ngữ cảnh $c_1 = (\text{“Weekend”, “Home”, “Kids”})$, $c_2 = (\text{“Weekday”, “Home”, “Family”})$, $c_3 = (\text{“Weekend”, “Cinema”, “Partner”})$. Mục đích của việc chuyển đổi $r_{ixt} \in [0,1]$ để sử dụng trong phương pháp tính toán mức độ phù hợp giữa người dùng với sản phẩm trong một tình huống ngữ cảnh. Nội dung này sẽ được trình bày chi tiết trong các mục tiếp theo của bài báo.

Bảng 2. Ma trận đánh giá chuyển đổi của lọc cộng tác theo ngữ cảnh

User	Item	Rating	Context Situation
u_1	p_1	1	c_1
u_1	p_1	0,8	c_2
u_2	p_1	0,6	c_3
u_2	p_1	0,8	c_2
u_3	p_3	0,6	c_3
u_3	p_2	0,4	c_3

Trên cơ sở ma trận đánh giá xác định theo (3), chúng tôi tiếp cận huấn luyện dữ liệu này trên cơ sở mô hình đồ thị $G=\langle V,E \rangle$. Tập đỉnh V của đồ thị được chia thành ba tập: tập người dùng, tập sản phẩm và tập tình huống ngữ cảnh ($V=U \cup P \cup C$). Tập cạnh E của đồ thị được xác định theo công thức (4). Mỗi cạnh $e \in E$ đều có dạng $e = (i, t)$ hoặc $e = (t, x)$, trong đó $i \in U, x \in P$ và $t \in C$. Trọng số của mỗi cạnh được xác định theo (5), (6).

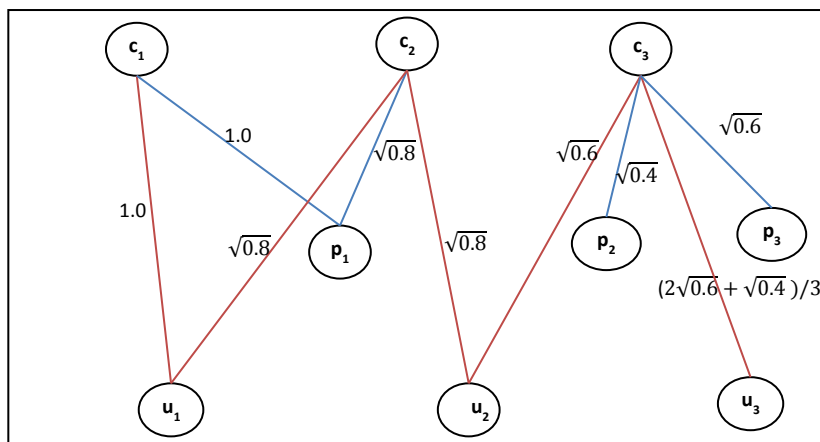
$$E = \{(i, t) \cup e = (t, x) \mid r_{it} \neq \phi \wedge r_{tx} \neq \phi\} \tag{4}$$

$$w_{it} = \begin{cases} r_{it} & \text{if } (i, t) \in E \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$w_{tx} = \begin{cases} r_{tx} & \text{if } (t, x) \in E \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Từ ma trận đánh giá đa chiều ban đầu, theo (3) ta xác định được r_{ixt} . Biểu diễn trên đồ thị r_{ixt} chính là trọng số đường đi có độ dài 2 từ người dùng i qua tình huống ngữ cảnh t tới sản phẩm x . Trọng số đường đi có độ dài 2 từ người dùng i qua tình huống ngữ cảnh t tới sản phẩm x này sẽ được tính bằng tích trọng số của các cạnh tương ứng. Hai cạnh tương ứng ở đây được biết đến là cạnh $e = (i, t)$ và cạnh $e = (t, x)$. Chúng tôi quy ước trọng số cạnh $e = (i, t)$ là r_{it} và trọng số cạnh $e = (t, x)$ là r_{tx} , khi đó: $r_{it} * r_{tx} = r_{ixt}$. Vì vậy chúng tôi sẽ coi $r_{it} = r_{tx} = \sqrt{r_{ixt}}$ để đảm bảo cân bằng trọng số đường đi từ đỉnh người dùng tới đỉnh tình huống ngữ cảnh và đỉnh tình huống ngữ cảnh tới đỉnh sản phẩm. Trong trường hợp có hơn 1 đường đi từ 1 đỉnh người dùng tới 1 đỉnh tình huống ngữ cảnh, chúng tôi sẽ tính trọng số đường đi từ đỉnh người dùng tới đỉnh tình huống ngữ cảnh này bằng cách lấy trung bình trọng số các cạnh đó. Cách tính tương tự với trọng số đường đi từ đỉnh tình huống ngữ cảnh tới đỉnh sản phẩm.

Ví dụ với hệ lọc cộng tác theo ngữ cảnh cho trong Bảng 2 thì mô hình đồ thị $G = \langle V, E \rangle$ được đưa ra như Hình 1.



Hình 1. Đồ thị biểu diễn cho lọc cộng tác theo ngữ cảnh

Trên cơ sở biểu diễn đồ thị cho lọc cộng tác theo ngữ cảnh, chúng tôi đề xuất phương pháp tính mức độ phù hợp của sản phẩm với người dùng trong tình huống ngữ cảnh cụ thể. Phương pháp đề xuất cho phép khai thác tất cả những mối quan hệ trực tiếp và gián tiếp từ người dùng i tới sản phẩm p trong tình huống ngữ cảnh c trên mô hình đồ thị để giải quyết vấn đề thừa dữ liệu và nâng cao hiệu quả dự đoán của hệ tư vấn cộng tác theo ngữ cảnh. Nội dung cụ thể của phương pháp dự đoán đề xuất sẽ được trình bày trong phần 3.2 sau.

3.2. Mức độ phù hợp của sản phẩm với người dùng trong tình huống ngữ cảnh cụ thể

Về cơ bản, các phương pháp tư vấn cộng tác truyền thống sinh ra tư vấn các sản phẩm cho người dùng dựa trên mức độ tương tự giữa các cặp người dùng hoặc mức độ tương tự giữa các cặp sản phẩm. Ví dụ theo Bảng 2, để đưa ra tư vấn cho người dùng u_1 , các phương pháp tư vấn cộng tác dựa vào người dùng sẽ tính toán mức độ tương tự giữa người dùng u_1 với các người dùng khác trong hệ thống (u_2, u_3). Chúng ta thấy rằng, mức độ tương tự giữa người dùng u_1 và u_2 có thể tính được do 2 người dùng này đều đưa ra đánh giá cho sản phẩm p_1 trong các tình huống ngữ cảnh khác nhau, nên p_2 được đánh giá bởi người dùng u_2 sẽ được tư vấn cho người dùng u_1 . Nhưng mức độ tương tự giữa người dùng u_1 và u_3 dựa vào các độ tương quan hoặc độ đo tương tự không tính được vì 2 người dùng này không cùng đánh giá chung sản phẩm nào của hệ thống. Do vậy p_3 được đánh giá bởi u_3 sẽ không được dùng để tư vấn cho u_1 . Vấn đề tương tự gặp phải với phương pháp tư vấn cộng tác dựa vào sản phẩm.

Tuy nhiên khi quan sát trên mô hình đồ thị trong Hình 1 thì mức độ phù hợp giữa người dùng u_1 và sản phẩm p_3 sẽ được xác định bằng tổng trọng số của các đường đi từ đỉnh u_1 tới đỉnh p_3 . Tổng trọng số các đường đi từ u_1 tới p_3 thu được bằng cách quan sát tất cả các đường đi từ u_1 tới p_3 , có 2 đường đi như vậy là $u_1 - c_2 - u_2 - c_3 - p_3$ và $u_1 - c_1 - p_1 - c_2 - u_2 - c_3 - p_3$. Trong đó, trọng số của mỗi đường đi sẽ được tính bằng tích trọng số của các cạnh tương ứng trên đường đi đó. Do vậy, đường đi $u_1 - c_2 - u_2 - c_3 - p_3$ (độ dài 4) có trọng số là $\sqrt{0,8} \times \sqrt{0,8} \times \sqrt{0,6} \times \sqrt{0,6} = 0,48$, đường đi $u_1 - c_1 - p_1 - c_2 - u_2 - c_3 - p_3$ (độ dài 6) có trọng số là $1 \times 1 \times \sqrt{0,8} \times \sqrt{0,8} \times \sqrt{0,6} \times \sqrt{0,6} = 0,48$ nên mức độ phù hợp giữa u_1 và p_3 là: $0,48 + 0,48 = 0,96$. Như vậy có thể thấy nếu ta mở rộng độ dài đường đi từ đỉnh người dùng tới đỉnh sản phẩm thì sẽ khai thác được nhiều mối quan hệ bắc cầu giữa người dùng và sản phẩm, góp phần giải quyết vấn đề thưa dữ liệu và tăng độ chính xác dự đoán trong hệ tư vấn cộng tác theo ngữ cảnh.

Tổng quát, mức độ phù hợp của người dùng u_i với sản phẩm p_x trong điều kiện ngữ cảnh c_t trên đồ thị được ước lượng bằng tổng các trọng số của tất cả các đường đi có độ dài chẵn $L (L=2,4,6, \dots)$ từ u_i qua c_t tới p_x . Trọng số của mỗi đường đi sẽ được tính bằng tích trọng số của các cạnh tương ứng trên đường đi đó. Khi đó, bài toán dự đoán đánh giá cho hệ tư vấn cộng tác theo ngữ cảnh chính là bài toán tìm kiếm đường đi từ người dùng u_i qua c_t tới p_x trên mô hình đồ thị cho trước. Bằng cách tiếp cận này, mức độ phù hợp giữa người dùng với sản phẩm trong tình huống ngữ cảnh cụ thể sẽ được xác định dựa trên việc khai thác tất cả các mối quan hệ giữa người dùng với sản phẩm trên đồ thị một cách trực tiếp và gián tiếp từ đỉnh người dùng tới đỉnh sản phẩm mà đi qua tình huống ngữ cảnh xác định trên đồ thị G.

Với dữ liệu đầu vào là ma trận $W(N \times K)$ được xác định theo công thức (5) thể hiện mối quan hệ giữa người dùng với tình huống ngữ cảnh, ma trận $W(K \times M)$ được xác định theo công thức (6) thể hiện mối quan hệ giữa tình huống ngữ cảnh và sản phẩm. Khi đó quá trình tính mức độ phù hợp của người dùng với sản phẩm trên đồ thị có thể hiện toán học chính là phép nhân ma trận theo công thức (7).

$$W^L \langle N \times M \rangle \cong \begin{cases} W \langle N \times K \rangle \times W \langle K \times M \rangle & \text{if } L=2 \\ W \langle N \times K \rangle \times W^T \langle N \times K \rangle \times W^{L-2} \langle N \times M \rangle & \text{if } L=4, 6, 8, \dots \end{cases} \quad (7)$$

Ma trận $W(N \times M)$ biểu diễn mối quan hệ giữa người dùng với sản phẩm được xác định theo (7) cho biết mức độ phù hợp của người dùng với sản phẩm. Mức độ phù hợp của người dùng $i \in U$ đối với sản phẩm mới $x \in P$ phụ thuộc vào độ dài đường đi L từ đỉnh người dùng 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 $i \in U$ và $x \in P$.

Khi đó thể hiện trên đồ thị thì chúng ta đã thiết lập được các đường đi trực tiếp từ người dùng u_i tới sản phẩm p_x với trọng số chính là mức độ phù hợp của sản phẩm với người dùng đó xác định theo (7). Vì mức độ phù hợp của sản phẩm p_x với người dùng u_i xác định được tới giai đoạn này chưa tích hợp thông tin ngữ cảnh, do vậy để xác định mức độ phù hợp của người dùng u_i với sản phẩm p_x trong điều kiện ngữ cảnh c_t thì chúng tôi sẽ tiến hành mở rộng đường đi từ u_i qua p_x tới c_t . Trọng số đường đi từ u_i qua p_x tới c_t sẽ được tính bằng tích trọng số các cạnh tương ứng, được xác định theo công thức (8).

$$r_{ixt} = w_{ix} \times w_{xt} \quad (8)$$

Theo công thức (7), (8) chúng ta đã xác định mức độ phù hợp của người dùng u_i với sản phẩm p_x trong điều kiện ngữ cảnh c_t . Để tư vấn các sản phẩm mới phù hợp với mỗi người dùng u_i trong điều kiện ngữ cảnh c_t , chúng tôi sẽ chọn ra K sản phẩm p_x thỏa mãn r_{ixt} lớn nhất, tức có mức độ phù hợp cao nhất với người dùng trong điều kiện ngữ cảnh đã cho.

Trên cơ sở các công thức đưa ra ở trên, chúng tôi đề xuất thuật toán tư vấn cộng tác theo ngữ cảnh dựa trên mô hình đồ thị (CACF_Graph) được miêu tả chi tiết trong Hình 2.

Đầu vào:

- Ma trận đánh giá đa chiều cho lọc cộng tác theo ngữ cảnh.
- $i \in U$ là người dùng cần được tư vấn (người dùng hiện thời).
- $t \in (C_1 \times \dots \times C_n)$ là tình huống ngữ cảnh của người dùng hiện thời.
- L là độ dài đường đi từ người dùng hiện thời tới các sản phẩm của hệ thống.
- K là số lượng sản phẩm cần tư vấn cho người dùng hiện thời.

Đầu ra:

- Tư vấn K sản phẩm $x \in P$ có mức độ phù hợp cao nhất cho người dùng hiện thời $i \in U$ trong tình huống ngữ cảnh $t \in C$

Các bước tiến hành:

Bước 1. Khởi tạo các ma trận $W(N \times K)$, $W(K \times M)$ theo công thức (5), (6).

Bước 2. Tính toán mức độ phù hợp giữa người dùng và sản phẩm

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

Repeat

$$W^L \langle N \times M \rangle \approx \begin{cases} W \langle N \times K \rangle \times W \langle K \times M \rangle & \text{if } L=2 \\ W \langle N \times K \rangle \times W^T \langle N \times K \rangle \times W^{L-2} \langle N \times M \rangle & \text{if } L=4, 6, 8, \dots \end{cases}$$

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

Until ($w_{ix}^L \neq 0$ với mọi $i \in U, x \in P$)

Bước 3. Tính toán mức độ phù hợp của người dùng $i \in U$ với các sản phẩm $x \in P$ trong điều kiện ngữ cảnh $t \in C$ theo công thức (8):

$$r_{ixt} = w_{ix} \times w_{xt}$$

Bước 4. Sắp xếp r_{ixt} theo thứ tự giảm dần.

Bước 5. Chọn K sản phẩm mới $x \in P$ đầu tiên tư vấn cho người dùng $i \in U$

Hình 2. Thuật toán CACF_GRAPH

Một trong những khó khăn khi thực hiện thuật toán CACF_GRAPH dựa trên đồ thị ở trên là tại bước 2 chúng ta phải thực hiện nhiều phép nhân ma trận để tính toán $W^L \langle N \times M \rangle$, điều này khó có thể thực hiện được khi số lượng người dùng, sản phẩm, tình huống ngữ cảnh và giá trị L lớn. Để giải quyết vấn đề này, chúng tôi tiến hành áp dụng thuật toán kích hoạt lan truyền trên đồ thị để tính toán mức độ phù hợp của người dùng với sản phẩm trong tình huống ngữ cảnh cụ thể. Có một số thuật toán kích hoạt lan truyền trên đồ thị được biết đến như Constrained Leaky Capacitor Model (LCM), Branch-and-Bound, Hopfiled Net, Google PageRank [23]. Sử dụng các thuật toán này để tìm đường đi trên đồ thị phục vụ cho bước 2 sẽ giảm thời gian huấn luyện trong các hệ tư vấn thực tế. Vì vậy trong bước 2 của thuật toán đề xuất CACF_GRAPH, chúng tôi sẽ áp dụng giải thuật kích hoạt lan truyền trên đồ thị để nâng cao hiệu suất huấn luyện mô hình tư vấn.

Một trong số các thuật toán lan truyền kích hoạt phổ biến hiện nay là Google PageRank [7]. Google PageRank được áp dụng thành công trong các máy tìm kiếm của Google. Ý tưởng của thuật toán này là khai thác các mối quan hệ giữa các thực thể của hệ thống trên một đồ thị có trọng số [7]. Chúng tôi sẽ áp dụng thuật toán kích hoạt lan truyền Google PageRank cho bước 2 của CACF_GRAPH. Thuật toán được miêu tả như sau:

Giả sử, $u_a \in U$ là người dùng hiện thời cần được tư vấn. Tại bước 2 của CACF_GRAPH chúng tôi cần tính mức độ phù hợp của người dùng u_a với các sản phẩm trong hệ thống dựa vào Google PageRank. Ký hiệu N_a là tập những đỉnh mà u_a có thể kết nối tới, $n_i \in N_a$ có thể coi là 1 đỉnh tình huống ngữ cảnh hoặc 1 đỉnh sản phẩm. Ký hiệu e_{ij} là trọng số kết nối giữa 2 đỉnh n_i và n_j , e_{aj} là trọng số kết nối giữa 2 đỉnh u_a và n_j . Các giá trị e_{ij} được khởi tạo từ ma trận $W(N \times K)$, $W(K \times M)$ được xác định theo công thức (5), (6). Gọi $a_i(t)$ là trọng số của đường đi kết nối từ đỉnh người dùng u_a tới đỉnh n_i khi khai thác các đường đi có độ dài L . Thuật toán sẽ dừng lại sau L bước lặp. Thuật toán được miêu tả trong Hình 3 sau.

Đầu vào:

- Ma trận (e_{ij}) được khởi tạo từ ma trận $W(N \times K)$, $W(K \times M)$ được xác định theo công thức (3), (4).
- α là hằng số điều chỉnh trọng số đường đi ($0 \leq \alpha \leq 1$).
- $u_a \in U$ là người dùng cần được tư vấn (người dùng hiện thời).

Đầu ra:

- Mức độ phù hợp của người dùng u_a với sản phẩm $p_x \in P$.

Các bước tiến hành:

1. Thiết lập $a_i(0) = w_{ii}$ nếu n_i là đỉnh tình huống ngữ cảnh, $a_i(0) = w_{ix}$ nếu n_i là đỉnh sản phẩm (w_{ii} được khởi tạo từ ma trận $W(N \times K)$, w_{ix} được khởi tạo từ ma trận $W(K \times M)$).
2. for $t = 0, 1, 2, \dots, L$
3. for $n_i \in N_a$ do
4. $a_i(t) = e_{ai}$;

```

5.   for  $n_j \in N_a$  do
6.      $a_i(t) \leftarrow a_i(t) + \alpha \cdot e_{ij} \cdot a_j(t-1)$ ;
7.     if ( $e_{ij} > 0$  and  $u_j \neq p_x$ ) then
8.       return ( $a_i(L)$ ) ;
9.     else if ( $t=L$ ) break;
10.  endfor
11. endfor
12. endfor
13. return ( $a_i(L)$ ); // Trọng số của đường đi từ đỉnh người dùng  $u_a$  tới đỉnh sản phẩm  $p_x$  có độ dài  $L$ .

```

Hình 3. Thuật toán Google PageRank tính mức độ phù hợp của người dùng với sản phẩm

IV. THỬ NGHIỆM VÀ ĐÁNH GIÁ

4.1. Dữ liệu thử nghiệm

Để thấy rõ hiệu quả của phương pháp đề xuất, chúng tôi thực hiện tiến hành thử nghiệm trên ba bộ dữ liệu DepaulMovie [26], MovieLens [27], InCarMusic[28].

- 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 MovieLens 100K chứa 100000 đánh giá từ 973 người dùng, 1682 phim trong các tình huống ngữ cảnh khác nhau. Bộ dữ liệu này có 2 chiều ngữ cảnh là *TimeOfDay*, *TimeOfWeek*. Chiều ngữ cảnh *TimeOfDay* có 5 điều kiện ngữ cảnh (“Morning”, “Noon”, “Afternoon”, “Evening”, “Night”), chiều ngữ cảnh *TimeOfWeek* có 2 điều kiện ngữ cảnh (“Weekend”, “Weekday”). Các mức đánh giá nằm trong dải từ 1 đến 5, mức độ thưa thớt của dữ liệu là 93,89 %. 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.

4.2. Độ đo đánh giá

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. Căn cứ theo hai nhiệm vụ đó thì có hai nhóm độ đo đánh giá hệ thống tư vấn tương ứng là: Nhóm độ đo đánh giá hệ thống tư vấn cho bài toán dự đoán đánh giá và nhóm độ đo đánh giá cho bài toán tư vấn Top-N sản phẩm. Chúng tôi tập trung vào nhóm độ đo đánh giá cho bài toán tư vấn Top-N sản phẩm để đánh giá hiệu quả của phương pháp đề xuất trong sự so sánh với các phương pháp tư vấn theo ngữ cảnh cơ sở. Một số độ đo phổ biến được biết tới đó là Precision@N, Recall@N and F-measure@N [6] được miêu tả 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):

$$\text{Precision@N} = \frac{|{\text{relevant items}} \cap {\text{top-N items}}|}{N} \quad (9)$$

- 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):

$$\text{Recall@N} = \frac{|{\text{relevant items}} \cap {\text{top-N items}}|}{|{\text{relevant items}}|} \quad (10)$$

- 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 (11). Giá trị F-measure lớn thể hiện thuật toán tư vấn có độ chính xác càng cao:

$$F\text{-measure@N} = \frac{2}{\frac{1}{\text{Precision@N}} + \frac{1}{\text{Recall@N}}} \quad (11)$$

4.3. Phương pháp thử nghiệm

Để đánh giá độ chính xác của hệ thống tư vấn, trước tiên từ ma trận đánh giá R ta tiến hành chia các người dùng (các hàng trong ma trận R) thành hai phần, một phần U_{train} được sử dụng làm dữ liệu huấn luyện, phần còn lại U_{test} được sử dụng để kiểm tra sao cho $U_{train} \cup U_{test} = U$. Việc phân chia này đảm bảo 80 % người dùng thuộc tập U_{train} , 20 % người dùng còn lại thuộc tập U_{test} . Có một số cách phân chia tập dữ liệu U thành 2 tập U_{train} và U_{test} , đó là: Splitting, Bootstrap sampling, k-fold cross-validation [10]. Trong các cách phân chia này thì k-fold cross-validation là phương pháp được sử dụng rộng rãi và cho kết quả 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} . Trong bài báo này, chúng tôi sẽ sử dụng phương pháp k-fold cross-validation ($k = 10$) để tiến hành chia dữ liệu kiểm nghiệm.

Để đánh giá độ chính xác của hệ thống tư vấn hiện thời, ta tiến hành như sau: 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.

4.4. Kết quả thử nghiệm

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.
- *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.
- *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_ICS* (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.
- *ItemSplitting_Graph*: Vì phương pháp của Huang [23] chỉ áp dụng cho các hệ tư vấn cộng tác với 2 thông tin đầu vào là người dùng và sản phẩm, do vậy chúng tôi tiến hành sử dụng phương pháp phân chia sản phẩm theo ngữ cảnh để đưa ma trận đánh giá đa chiều theo ngữ cảnh về ma trận đánh giá 2 chiều (users, items). Ma trận đánh giá 2 chiều mới nhận được sẽ được dùng để huấn luyện và đưa ra tư vấn sử dụng phương pháp của Huang [23].

Bảng 3. Kết quả kiểm nghiệm của các phương pháp lọc cộng tác theo ngữ cảnh cho bộ dữ liệu DepaulMovie

Methods	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
UserSplitting-BiasedMF	0,089	0,327	0,140
ItemSplitting-BiasedMF	0,085	0,316	0,134

UISplitting-BiasedMF	0,081	0,298	0,127
CSLIM_C	0,063	0,214	0,097
CSLIM_ICS	0,065	0,218	0,100
CSLIM_MCS	0,085	0,315	0,134
ItemSplitting_Graph	0,117	0,324	0,172
CACF_GRAPH	0,120	0,343	0,178

Bảng 4. Kết quả kiểm nghiệm của các phương pháp lọc cộng tác theo ngữ cảnh cho bộ dữ liệu MovieLense 100K

Methods	Precision@10	Recall@10	F-measure@10
CPTF	0,022	0,016	0,0185
CAMF_C	0,028	0,020	0,0233
CAMF_CU	0,028	0,019	0,0226
CAMF_CI	0,024	0,015	0,0184
UserSplitting-BiasedMF	0,057	0,038	0,0456
ItemSplitting-BiasedMF	0,051	0,033	0,0401
UISplitting-BiasedMF	0,052	0,033	0,0404
CSLIM_C	0,0035	0,0024	0,0028
CSLIM_ICS	0,0037	0,0028	0,0032
CSLIM_MCS	0,0044	0,0034	0,0038
ItemSplitting_Graph	0,081	0,043	0,0562
CACF_GRAPH	0,092	0,051	0,0656

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

Methods	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
UserSplitting-BiasedMF	0,031	0,215	0,0542
ItemSplitting-BiasedMF	0,033	0,226	0,0576
UISplitting-BiasedMF	0,032	0,219	0,0558
CSLIM_C	0,018	0,127	0,0315
CSLIM_ICS	0,039	0,231	0,0667
CSLIM_MCS	0,018	0,106	0,0308
ItemSplitting_Graph	0,035	0,342	0,0635
CACF_GRAPH	0,043	0,364	0,0769

Kết quả kiểm nghiệm được thể hiện trong Bảng 3, Bảng 4, Bảng 5 cho thấy phương pháp đề xuất CACF_GRAPH cho lại độ chính xác tốt hơn so với các phương pháp lọc cộng tác theo ngữ cảnh cơ sở. Phương pháp đề xuất cũng cho độ chính xác tốt hơn so với phương pháp đưa dữ liệu đánh giá đa chiều về dữ liệu đánh giá 2 chiều đề áp dụng trực tiếp phương pháp của Huang trên đồ thị (ItemSplitting_Graph). Điều đó có thể khẳng định phương pháp đề xuất cho phép khai thác tất cả những mối quan hệ trực tiếp và gián tiếp từ người dùng tới sản phẩm trong các tình huống ngữ cảnh trên mô hình đồ thị để giải quyết vấn đề thưa dữ liệu và nâng cao hiệu quả dự đoán của hệ tư vấn cộng tác theo ngữ cảnh.

V. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Bài báo đã trình bày một phương pháp tiếp cận cho lọc cộng tác theo ngữ cảnh bằng mô hình đồ thị. Trong đó, phương pháp biểu diễn đồ thị đề xuất phù hợp với tất cả các bộ dữ liệu theo ngữ cảnh cho hệ tư vấn cộng tác theo ngữ cảnh hiện nay. Dựa vào biểu diễn này, chúng tôi đưa ra đề xuất phương pháp dự đoán các sản phẩm phù hợp với người dùng trong điều kiện ngữ cảnh cụ thể (CARS_GRAPH). Phương pháp dự đoán trên mô hình đồ thị này về bản chất được quy về bài toán tìm kiếm đường đi trên biểu diễn đồ thị đã cho. Kết quả kiểm nghiệm cho thấy, phương pháp đề xuất cho lại kết quả dự đoán tốt hơn các phương pháp lọc cộng tác theo ngữ cảnh cơ sở trong trường hợp các bộ dữ liệu có mức độ thưa thớt khác nhau. Điều đó có thể khẳng định, phương pháp tiếp cận cho lọc cộng tác theo ngữ cảnh dựa trên mô hình đồ thị cho phép khai thác tất cả những mối quan hệ trực tiếp và gián tiếp từ người dùng tới sản phẩm trong các tình huống ngữ cảnh để giải quyết vấn đề thưa dữ liệu và nâng cao hiệu quả dự đoán của hệ tư vấn cộng tác theo ngữ cảnh.

Trong thời gian tới, chúng tôi dự định sẽ mở rộng nghiên cứu của mình cho hệ tư vấn lai theo ngữ cảnh nhằm tích hợp được nhiều thông tin phục vụ cho quá trình huấn luyện nâng cao chất lượng tư vấn. Ngoài ra chúng tôi cũng

có kế hoạch nghiên cứu về việc tính mức độ tương tự giữa các chiều ngữ cảnh để nâng cao hiệu quả của phương pháp mô hình hóa ngữ cảnh phụ thuộc áp dụng cho hệ tư vấn theo ngữ cảnh.

TÀI LIỆU THAM KHẢO

- [1] G. Adomavicius, A. Tuzhilin. "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 (2005).
- [2] G. Adomavicius, B. Mobasher, F. Ricci, and A. Tuzhilin. Context-ware recommender systems. *AI Managine*, vol. 32, no.3, pp.67-80, 2011.
- [3] Y. Zheng, R. Burke, B. Mobasher. Differential context relaxation for context-aware travel recommendation. *In: 13th International Conference on Electronic Commerce and Web Technologies (EC-WEB 2012)*, 2012, pp.88-89.
- [4] C. Ono, Y. Takinshima, Y. Motomura, H. Asoh. Context-aware preference model based on a study of difference between real and supposed situation data. *User Modeling, Adaptation, and Personalization*(2009) 102-113.
- [5] G. Adomavicius, R. Sankaranarayanan, S. Sen, A.Tuzhilin, Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems (TOIS) 23 (1) (2005)* 103-145.
- [6] 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.
- [7] Jason Weston, Andre Elisseeff, Dengyong Zhou, Christina S. Leslie, and William Stafford Noble. Protein ranking: From local to structure in the protein similarity network. *Proceedings of the National Academy of Sciences of the United States of America*. April 27, 2004. 101 (17) 6559-6563
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [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 4th ACM Recommender Systems Conference (RecSys '10)*, pp. 79-86, ACM, September 2010.
- [13] B. Zou, C. Li, L. Tan, and H. Chen. GPU-TENSOR: efficient tensor factorization for context-aware recommendations. *Information Sciences*, vol. 299, pp. 159-177, 2015.
- [14] L Baltrunas, B Ludwig, F Ricci. Matrix factorization techniques for context aware recommendation. *ACM RecSys*, 2011.
- [15] Y Zheng, B Mobasher, R Burke. CSLIM: Contextual SLIM Recommendation Algorithms. *ACM RecSys*, 2014.
- [16] Y Zheng, B Mobasher, R Burke. Deviation-Based Contextual SLIM Recommenders. *ACM CIKM*, 2014.
- [17] 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.
- [18] Şamdan, Emrah & Taşçı, Arda & Cicekli, Nihan. (2014). A Graph-based Collaborative and Context-aware Recommendation system for TV programs. *RecSys 2014 TV Workshop*.
- [19] 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.
- [20] 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.
- [21] 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.

- [22] 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.
- [23] Z. Huang, D. Zeng, H. Chen. “Analyzing Consumer-product Graphs: Empirical Findings and Applications in Recommender Systems”. *Management Science*, 53(7), 1146-1164 (2007).
- [24] Do Thi Lien, Nguyen Duy Phuong. “*Collaborative Filtering with a Graph-based Similarity Measure*”. CommandTel, 2014.
- [25] 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)*.
- [26] Y. Zheng, B. Mobasher và R. Burke. CARSKit: A Java-Based Context-Aware Recommendation Engine, 2015.
- [27] <http://www.grouplens.org/>.
- [28] https://github.com/irecsys/CARSKit/blob/master/context-aware_data_sets/Music_InCarMusic.zip.

A NEW CONTEXT-AWARE COLLABORATIVE FILTERING METHOD

Do Thi Lien, Nguyen Duy Phuong

ABSTRACT: Collaborative filtering is one of the most popular method in building recommender systems. The recent collaborative filtering methods explore only rating datasets between users with items to find group of users having similar interest in items. From these information, the system can recommend suitable items for active users. In this paper, we propose a new Context-Aware Collaborative Filtering (CACF) method. The proposed method allows represent relationships between users, items and context situations on an unified graph model. Based on the graph representation for CACF, we then build a new model to give predictions and make item recommendations for the target user in a specific context situation. Experimental results on some real datasets show that the proposed method outperforms more significant than baseline methods of CACF.

Keywords: Context-aware recommender system - CARS; Context-aware collaborative filtering - CACF; Context; Graph-based contextual modelling.