

MỘT PHƯƠNG PHÁP MÔ HÌNH HÓA NGỮ CẢNH CHO HỆ TƯ VẤN

Đỗ Thị Liên

Học viện Công nghệ Bưu chính Viễn thông

Tóm tắt: Hệ tư vấn 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. Trong các cách tiếp cận xây dựng hệ tư vấn thì cách tiếp cận dựa trên ngữ cảnh được chứng minh là cải thiện đáng kể chất lượng tư vấn trong nhiều ứng dụng thực tế so với hệ tư vấn theo tiếp cận truyền thống trước đây. Tuy nhiên, lĩnh vực này còn tồn tại nhiều thách thức, trong đó khó khăn điển hình là làm thế nào để tích hợp thông tin ngữ cảnh và vấn đề dữ liệu thưa, điều này ảnh hưởng trực tiếp tới chất lượng tư vấn. Trong bài báo này, tác giả đề xuất một phương pháp mô hình hóa ngữ cảnh cho hệ tư vấn cho phép tích hợp đầy đủ thông tin ngữ cảnh và giải quyết hiệu quả vấn đề dữ liệu thưa. Trong đó, mối quan hệ giữa các người dùng, sản phẩm và ngữ cảnh được biểu diễn trên cùng một mô hình đồ thị 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 đối tượng dữ liệu trong quá trình huấn luyện dựa trên mô hình đồ thị cho lọc cộng tác. 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: 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 (Contextual modeling).

I. MỞ ĐẦU

Với sự bùng nổ của thông tin trên Internet thì cần thiết phải có công cụ giúp người dùng lựa chọn các thông tin trực tuyến phù hợp với mình. Để đáp ứng nhu cầu này, các hệ thống tư vấn đã ra đời, đây là lĩnh vực này đang có sự phát triển rất nhanh và mạnh. Các nghiên cứu về việc xây dựng hệ tư vấn truyền thống trước đây chủ yếu tập trung khai thác mối quan hệ giữa hai nhóm đối tượng là người dùng (user) và sản phẩm (item) để thực hiện huấn luyện và đưa ra gợi ý. Bài toán tư vấn truyền thống có thể được biểu diễn dựa trên ma trận đánh giá hai chiều sau:

$$R_o: User \times Item \rightarrow Rating \quad (1)$$

Trên thực tế, đánh giá của người dùng không cố định, mà thay đổi theo ngữ cảnh bên ngoài. Ví dụ một người trời nóng thì thích uống sinh tố, nhưng khi trời lạnh lại thích

uống cà phê nóng. Hoặc cùng một bộ phim nhưng trời mưa thì thích xem còn trời khô ráo thì 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 yêu thích [4], tư vấn phim và âm nhạc [3], 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 bạn đồng hành (một mình, gia đình, bạn bè), thời gian (buổi trong ngày, thời gian trong tuần, mùa). Khi đó, bài toán tư vấn theo ngữ cảnh sẽ được biểu diễn dựa trên ma trận đánh giá đa chiều như sau:

$$R_1: User \times Item \times Context \rightarrow Rating \quad (2)$$

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 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 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 qui 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

Tác giả liên hệ: Đỗ Thị Liên,

Email: liendt@ptit.edu.vn

Đến tòa soạn: 9/2021, chỉnh sửa: 10/2021, chấp nhận đăng: 10/2021.

kiến ngữ cảnh là giá trị của một chiều ngữ cảnh, ví dụ: 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 (“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, 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 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. Trong bài báo này, tác giả đề xuất một phương pháp mô hình hóa ngữ cảnh mới cho hệ tư vấn. 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. Trên cơ sở biểu diễn đồ thị đó, tác giả tiến hành tính toán mức độ tương tự giữa các cặp người dùng dựa vào quá trình tìm kiếm tất cả các đường đi từ đỉnh người dùng tới đỉnh người dùng đi qua đỉnh tình huống ngữ cảnh. Việc tính toán mức độ tương tự giữa các cặp sản phẩm dựa vào mô hình đồ thị đề xuất được thực hiện tương tự. Từ độ đo tương tự tính toán dựa trên mô hình đồ thị đề xuất, tác giả xây dựng thuật toán dự đoán mức độ phù hợp của sản phẩm với người dùng trong từng điều kiện ngữ cảnh cụ thể theo phương pháp k-láng giềng gần nhất.

Như vậy, việc phân tích lý thuyết chỉ ra rằng phương pháp mô hình hóa ngữ cảnh cho hệ tư vấn dựa trên mô hình đồ thị đề xuất cho phép tích hợp đầy đủ thông tin ngữ cảnh vào quá trình tư vấn và vấn đề dữ liệu thừa được giải quyết qua việc khai thác đầy đủ mối quan hệ trực tiếp và gián tiếp giữa các đối tượng dữ liệu trong quá trình huấn luyện. Bên cạnh đó, 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, 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

Adomavicius and Tuzhilin [1] là các nhà nghiên cứu đầu tiên tích hợp ngữ cảnh vào hệ tư vấn, 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, điển hình là tập trung vào cải tiến các phương pháp tư vấn theo ngữ cảnh. Các phương pháp này tiếp cận theo một trong ba hướng chính: 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, 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 dụng để huấn luyện và dự đoán bởi các phương pháp tư vấn truyền thống [2][6] [8][11]. Trái ngược với hướng lọc trước theo ngữ cảnh, lọc sau theo ngữ cảnh sẽ sử dụng toàn bộ ma trận đánh giá đã loại bỏ đi các chiều ngữ cảnh để

huấn luyện bởi các phương pháp tư vấn truyền thống, kết quả dự đoán thu được sẽ được lọc lại một lần nữa bởi ngữ cảnh hiện thời nhằm 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, đối với hai hướng tiếp cận lọc trước theo ngữ cảnh và lọc sau theo ngữ cảnh thì thông tin ngữ cảnh không được tích hợp trong quá trình huấn luyện.

Một hướng tiếp cận thứ ba cho phép tích hợp trực tiếp ngữ cảnh vào quá trình huấn luyện và tư vấn đó là mô hình hóa ngữ cảnh. Theo hướng tiếp cận thứ ba 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. 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. Hai kỹ thuật được sử dụng trong phương pháp mô hình hóa ngữ cảnh phụ thuộc là mô hình hóa dựa trên độ chênh lệch và mô hình hóa dựa trên độ tương tự [21]. 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 [9] và SLIM [10, 15] 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 [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 dựa trên các giải thuật tư vấn truyền thống là vấn đề dữ liệu thừa, độ phức tạp tính toán và khả năng mở rộng của nó.

Một hướng tiếp cận mô hình hóa ngữ cảnh 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 thừa và khả năng mở rộng dữ liệu [16]. Neves ARM và các cộng sự của mình

[19] 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[20] 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 [17] 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 [18] 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 hầu như chưa khái quát lên một giải pháp tổng thể chung cho các bài toán tư vấn theo ngữ cảnh.

Để giảm thiểu những hạn chế nêu trên, tác giả đề xuất một phương pháp mô hình hóa theo ngữ cảnh mới dựa trên mô hình đồ thị hợp nhất cho hệ tư vấn cộng tác để thực hiện huấn luyện và tư vấn các sản phẩm phù hợp với người dùng trong từng tình huống ngữ cảnh cụ thể. Phương pháp đề xuất được đánh giá là giải quyết việc tích hợp đầy đủ các thông tin ngữ cảnh và giải quyết hiệu quả vấn đề dữ liệu thưa của ma trận đánh giá đa chiều, từ đó nâng cao chất lượng của hệ tư vấn theo ngữ cảnh. Nội dung cụ thể của phương pháp đề xuất được trình bày chi tiết trong Mục III.

III. PHƯƠNG PHÁP ĐỀ XUẤT

Về cơ bản các phương pháp đề xuất đã có để giải quyết bài toán tư vấn cộng tác nói chung và theo ngữ cảnh nói riêng dựa trên mô hình đồ thị [1, 22, 23] đều tiến hành giải quyết ba vấn đề: *Phương pháp biểu diễn đồ thị cho bài toán tư vấn cộng tác theo ngữ cảnh, phương pháp tính toán mức độ tương tự giữa các cặp người dùng (hoặc các cặp sản phẩm) và xây dựng thuật toán dự đoán quan điểm của người dùng đối với các sản phẩm trong tình huống ngữ cảnh dựa vào đồ thị*. Trong bài báo này, tác giả mở rộng mô hình biểu diễn trên đồ thị đã được Huang [22] đề xuất cho hệ tư vấn cộng tác truyền thống sang cho hệ tư vấn cộng tác cho ngữ cảnh. Tiếp đến, tác giả đề xuất một phương pháp tính toán mức độ tương tự giữa các cặp người dùng dựa trên biểu diễn đồ thị nhằm giải quyết vấn đề dữ liệu thưa thớt. Độ tương tự giữa các cặp người dùng tính được sẽ là cơ sở để dự đoán đánh giá của người dùng với sản phẩm mới trong từng ngữ cảnh cụ thể theo phương pháp k-láng giềng gần nhất. Phương pháp lọc cộng tác theo ngữ cảnh dựa vào sản phẩm trên cơ sở mô hình đồ thị được thực hiện tương tự. Hai phương pháp đề xuất cho lại kết quả dự đoán tốt hơn so với các phương pháp tư vấn cộng tác theo ngữ cảnh cơ sở. Phương pháp đề xuất cũng nhằm đơn giản trong cài đặt để có thể dễ dàng trong việc triển khai thực tế.

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

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 $t \in C$.

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

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 [22]. 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 $c_1 = (\text{"Weekend"}, \text{"Home"}, \text{"Kids"})$, $c_2 = (\text{"Weekday"}, \text{"Home"}, \text{"Family"})$, $c_3 = (\text{"Weekend"}, \text{"Cinema"}, \text{"Partner"})$. Mục đích của việc chuyển đổi $r_{itx} \in [0,1]$ để sử dụng trong phương pháp tính toán mức độ tương tự giữa các cặp người dùng hoặc các cặp sản phẩm. Nội dung này sẽ được trình bày chi tiết trong các mục tiếp theo của bài báo.

Ví dụ chuyển đổi ma trận đánh giá đa chiều ban đầu cho trong Bảng 1 về dạng Bảng 2 như sau:

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

Từ ma trận đánh giá xác định theo (3), tác giả chuyển đổi biểu diễn mối quan hệ giữa các đối tượng dữ liệu 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, 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).

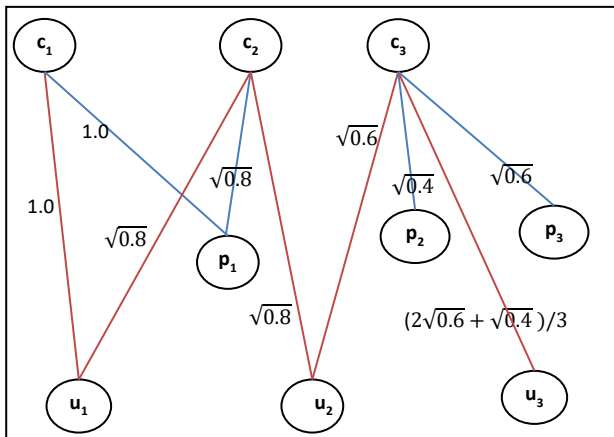
$$E = \{e = (i, t) \cup e = (t, x) | r_{it} \neq \emptyset \wedge r_{tx} \neq \emptyset\} \quad (4)$$

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

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

Từ ma trận đánh giá đa chiều ban đầu, theo (3) ta xác định được r_{itx} . Biểu diễn trên đồ thị r_{itx} 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 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 $e = (t, x)$. Tác giả qui ước trọng số cạnh $e = (i, t)$ là r_{it} và trọng số cạnh $e = (t, x)$ là r_{tx} , khi đó: $r_{itx} = r_{it} * r_{tx}$. Vì vậy tác giả coi $r_{it} = r_{tx} = \sqrt{r_{itx}}$ để đảm bảo cân bằng trọng số đường đi từ đỉnh người dùng tới đỉnh tình huống ngữ cảnh và từ 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, tác giả 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

B. Mức độ tương tự giữa các cặp người dùng cho lọc cộng tác theo ngữ cảnh dựa trên biểu diễn đồ thị

Để đưa ra tư vấn cho người dùng hiện thời, các phương pháp tư vấn cộng tác thường dựa trên mức độ tương tự giữa các cặp người dùng hoặc giữa các cặp sản phẩm. Việc tính toán mức độ tương tự này thường sử dụng các độ đo tương quan hoặc độ đo tương tự phổ biến như: Độ đo khoảng cách (Euclid, Minkowski...), độ đo tương tự (Cosin, Entropy...), độ đo tương quan (Pearson, Spearman Rank, Kendal,...). Tuy nhiên việc tính mức độ tương tự giữa các cặp người dùng thông qua các độ đo này trong nhiều trường hợp không thực hiện được khi các sản phẩm cả hai người dùng cùng đánh giá không giao nhau [1]. Nguyên nhân chính cho vấn đề này là do ma trận đánh giá quá thưa. 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 tương tự giữa 2 người dùng u_1 và u_3 sẽ được xác định bằng tổng trọng số các đường đi từ đỉnh u_1 tới đỉnh u_3 . Tổng trọng số các đường đi từ u_1 tới u_3 thu được bằng cách quan sát tất cả các đường đi từ u_1 tới u_3 , có 2 đường đi như vậy là $u_1 - c_2 - u_2 - c_3 - u_3$ và $u_1 - c_1 - p_1 - c_2 - u_2 - c_3 - u_3$. Trong đó, trọng số của mỗi đường đi sẽ được tính bằng tích trọng số của các cạnh tương ứng trên đường đi đó. Cụ thể, đường đi $u_1 - c_2 - u_2 - c_3 - u_3$ (độ dài 4) có trọng số là $\sqrt{0.8} * \sqrt{0.8} * \sqrt{0.6} * (2\sqrt{0.6} + \sqrt{0.4})/3 = 0.45$, đường đi $u_1 - c_1 - p_1 - c_2 - u_2 - c_3 - u_3$ (độ dài 6) có trọng số là $1 * 1 * \sqrt{0.8} * \sqrt{0.8} * \sqrt{0.6} * (2\sqrt{0.6} + \sqrt{0.4})/3 = 0.45$ nên mức độ tương tự giữa u_1 tới u_3 là: $0.45 + 0.45 = 0.9$. Như vậy có thể thấy nếu ta mở rộng độ dài đường đi từ đỉnh người dùng tới đỉnh người dùng thì sẽ khai thác được khá nhiều mối quan hệ bắc cầu giữa người dùng với người dùng, từ đó góp phần giải quyết hiệu quả vấn đề dữ liệu thưa của lọc cộng tác theo ngữ cảnh.

Tổng quát, mức độ tương tự giữa người dùng u_i và người dùng u_j 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 u_j . Trọng số của mỗi đường đi sẽ được tính bằng tích trọng số của các cạnh tương ứng trên đường đi đó. Khi đó, việc tính mức độ tương tự giữa các cặp người dùng cho lọc cộng tác theo ngữ cảnh chính là bài toán tìm kiếm đường đi từ đỉnh người dùng u_i qua c_t tới u_j trên mô hình đồ thị cho trước. Với cách tiếp cận này, mức độ tương tự giữa các cặp người dùng trong tình huống ngữ cảnh cụ thể sẽ được xác định dựa trên việc khai thác tất cả các mối quan hệ trực tiếp và gián tiếp giữa các cặp người dùng đó 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 tương tự giữa các cặp người dùng dựa trên đồ thị có thể hiện toán học chính là phép nhân ma trận theo công thức (7).

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

Ma trận $UZ(N \times N)$ biểu diễn mối quan hệ giữa các cặp người dùng được xác định theo (7) cho biết mức độ tương tự giữa các cặp người dùng đó. Mức độ tương tự này phụ thuộc vào độ dài đường đi L từ đỉnh người dùng đến đỉnh người dùng 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 $u_{ij}^L \neq 0$ với mọi $u_i \in U$ và $u_j \in U$.

Từ mức độ tương tự giữa các cặp người dùng xác định theo (7), với mỗi người dùng cần tư vấn u_a tác giả chọn k người dùng có mức độ tương tự cao nhất với u_a làm tập láng giềng. Kí hiệu U_a là tập láng giềng của u_a gồm k người dùng. Khi đó, đánh giá chưa biết của người dùng u_a với các sản phẩm p_x trong tình huống ngữ cảnh c_t theo công thức (8):

$$r_{axt} = \frac{\sum_{r_{jxt} \in R_j} r_{jxt}}{|R_j|}, R_j = \{r_{jxt} | r_{jxt} \neq 0, u_j \in U_a\} \quad (8)$$

Trên cơ sở các công thức đưa ra ở trên, tác giả đề xuất thuật toán tư vấn cộng tác theo ngữ cảnh dựa vào mức độ tương tự giữa các cặp người dùng trên mô hình đồ thị (CACF_UserBasedGraph) được miêu tả chi tiết trong Hình 2.

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 ngữ cảnh ứng với u_a
- L_{max} là giới hạn độ dài đường đi giữa các cặp người dùng
- K_1 là số lượng người dùng trong tập láng giềng với u_a
- K_2 là số lượng sản phẩm cần tư vấn cho u_a

Đầu ra:

- Danh sách K_2 sản phẩm tư vấn tới người dùng u_a trong tình huống ngữ cảnh c_t .

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 tương tự giữa các cặp người dùng

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

Repeat

$$UZ^L(N \times N) = \begin{cases} W(N \times K) * W(K \times N) & \text{if } L = 2 \\ W(N \times K) * W(K \times N) * UZ^{L-2}(N \times N) & \text{if } L = 4, 6, 8, \dots \end{cases}$$

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

Until $(uz_{ij}^L \neq 0 \text{ với mọi } u_j \in (U \setminus u_i) \text{ hoặc } L = L_{max})$;

Bước 3. Sinh tư vấn cho người dùng hiện thời u_a trong ngữ cảnh c_t

- Với mỗi người dùng hiện thời u_a , chọn K_1 người dùng có mức độ tương tự cao nhất với u_a làm tập láng giềng. Kí hiệu U_a là tập láng giềng của u_a gồm K_1 người dùng.
- Dự đoán đánh giá chưa biết r_{axt} của người dùng u_a với sản phẩm p_x trong tình huống ngữ cảnh c_t theo công thức (8).
- Chọn K_2 sản phẩm thực trong P có đánh giá dự đoán cao nhất để tư vấn cho người dùng u_a trong tình huống ngữ cảnh c_t .

Hình 2. Thuật toán CACF_UserBasedGraph

C. Mức độ tương tự giữa các cặp sản phẩm cho lọc cộng tác theo ngữ cảnh dựa trên biểu diễn đồ thị

Lập luận tương tự với mức độ tương tự giữa các cặp người dùng ở trên, tác giả đề xuất độ đo tương tự giữa các cặp sản phẩm dựa trên đồ thị có thể hiện toán học chính là phép nhân ma trận theo công thức (9)

$$PZ^L(M \times M) = \begin{cases} W(M \times K) * W(K \times M) & \text{if } L = 2 \\ W(M \times K) * W(K \times M) * PZ^{L-2}(M \times M) & \text{if } L = 4, 6, 8, \dots \end{cases} \quad (9)$$

Với mỗi sản phẩm p_x chưa được đánh giá bởi người dùng hiện thời u_a , gọi P_x là tập gồm K_1 sản phẩm tương tự nhất với p_x và đã được đánh giá bởi u_a . Khi đó, đánh giá chưa biết của người dùng u_a với sản phẩm p_x được dự đoán căn cứ vào trung bình đánh giá của người dùng u_a với các sản phẩm thuộc tập P_x .

$$r_{axt} = \frac{\sum_{r_{ajt} \in R_j} r_{ajt}}{|R_j|}, R_j = \{r_{ajt} | r_{ajt} \neq 0, p_j \in P_x\} \quad (10)$$

Trên cơ sở các công thức (9),(10) đưa ra, tác giả đề xuất thuật toán tư vấn cộng tác theo ngữ cảnh dựa vào mức độ tương tự giữa các cặp sản phẩm trên mô hình đồ thị (CACF_ItemBasedGraph) 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 ngữ cảnh ứng với u_a
- L_{max} là giới hạn độ dài đường đi giữa các cặp sản phẩm
- K_1 là số lượng sản phẩm trong tập láng giềng với sản phẩm được u_a đánh giá
- K_2 là số lượng sản phẩm cần tư vấn cho u_a

Đầu ra:

- Danh sách K_2 sản phẩm tư vấn tới người dùng u_a trong tình huống ngữ cảnh c_t .

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 tương tự giữa các cặp sản phẩm

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

Repeat

$$PZ^L(M \times M) = \begin{cases} W(M \times K) * W(K \times M) & \text{if } L = 2 \\ W(M \times K) * W(K \times M) * PZ^{L-2}(M \times M) & \text{if } L = 4, 6, 8, \dots \end{cases}$$

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

Until $(tz_{kj}^L \neq 0 \text{ với mọi } t_k \in (T \setminus t_j) \text{ hoặc } L = L_{max})$;

Bước 3. Sinh tư vấn cho người dùng hiện thời u_a trong ngữ cảnh c_t

- Thực hiện lặp: với mỗi sản phẩm $p_x \in P$ chưa được đánh giá bởi u_a
 - Chọn K_1 sản phẩm có mức độ tương tự cao nhất với p_x làm tập láng giềng. Kí

- hiệu P_x là tập láng giềng của p_x gồm K_1 sản phẩm.
- Dự đoán đánh giá chưa biết r_{ax_t} của người dùng u_a với sản phẩm p_x trong tình huống ngữ cảnh c_t theo công thức (10).
 - Chọn K_2 sản phẩm trong P có đánh giá dự đoán cao nhất để tư vấn cho người dùng u_a trong tình huống ngữ cảnh c_t .

Hình 3. Thuật toán CACF_ItemBasedGraph

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 thừa 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 thừa 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 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).

$$Precision@N = \frac{| \{relevant\ items\} \cap \{top-N\ items\} |}{N} \quad (11)$$

- 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)

$$Recall@N = \frac{| \{relevant\ items\} \cap \{top - N\ items\} |}{| \{relevant\ items\} |} \quad (12)$$

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

$$F\text{-measure@}N = \frac{2}{\frac{1}{Precision@N} + \frac{1}{Recall@N}} \quad (13)$$

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. 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.

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

Tác giả tiến hành kiểm nghiệm 2 thuật toán đề xuất (CACF_UserBasedGraph, CACF_ItemBasedGraph) trong sự so sánh với các phương pháp mô hình hóa cộng tác theo ngữ cảnh cơ sở sau:

- *CAMF_C* (Context-aware matrix factorization-Context) [9][14]: 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) [9][14]: 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) [9][14]: 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* [10][14]: 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) [10][14]: 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) [10][14]: Sử dụng phương pháp tuyến tính thưa SLIM kết hợp với mô hình hóa ngữ cảnh dựa trên độ tương quan đa chiều.

C. Kết quả thực nghiệm

Kết quả thực nghiệm được đưa ra trong Bảng 3, Bảng 4

Bảng 3. Giá trị Precision@10, Recall@10, F-measure@10 trên tập DepaulMovie

Phương pháp	Precision@10	Recall@10	F-measure@10
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
CACF_UserBasedGraph	0.104	0.335	0.159
CACF_ItemBasedGraph	0.109	0.338	0.165

Bảng 4. Giá trị Precision@10, Recall@10, F-measure@10 trên tập InCarMusic

Phương pháp	Precision@10	Recall@10	F-measure@10
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
CACF_UserBasedGraph	0.043	0.294	0.075
CACF_ItemBasedGraph	0.045	0.301	0.0783

Kết quả kiểm nghiệm được thể hiện trong Bảng 3, Bảng 4 cho thấy phương pháp đề xuất CACF_UserBasedGraph, CACF_ItemBasedGraph cho lại độ chính xác tốt hơn so với các phương pháp mô hình hóa cho lọc cộng tác theo ngữ cảnh cơ sở. Điều này được lý giải là vì mô hình đồ thị hợp nhất cho phép tích hợp đầy đủ người dùng, sản phẩm và tình huống ngữ cảnh, trên cơ sở đó tất cả các mối quan hệ trực tiếp và gián tiếp giữa các đỉnh được khai thác đầy đủ vào quá trình tính toán mức độ tương tự giữa các cặp người dùng hoặc các cặp sản phẩm. Đây chính là lợi thế trong việc nâng cao độ chính xác trong việc tính toán mức độ tương tự dựa trên mô hình đồ thị so với các phương pháp cơ sở, đặc biệt trong trường hợp dữ liệu thưa, từ đó nâng cao hiệu quả dự đoán đánh giá.

V. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Bài báo đã trình bày đề xuất một phương pháp mô hình hóa ngữ cảnh mới cho hệ tư vấn cộng tác theo ngữ cảnh. Phương pháp đề xuất cho phép tích hợp đầy đủ thông tin ngữ cảnh và giải quyết hiệu quả vấn đề dữ liệu thưa. Trong đó, việc tích hợp ngữ cảnh 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. Trên cơ sở biểu diễn đồ thị đó, tác giả tiến hành huấn luyện để đưa ra dự đoán đánh giá. Quá trình huấn luyện dựa trên mô hình đồ thị này được thực hiện qua 2 bước: 1) Tính toán mức độ tương tự giữa các cặp người dùng dựa vào quá trình tìm kiếm tất cả các đường đi từ đỉnh người dùng tới đỉnh người dùng. Như vậy tất cả các mối quan hệ trực tiếp và gián tiếp giữa các đỉnh trên mô hình đồ thị được khai thác triệt để trong việc tính toán mức độ tương tự giữa các cặp người dùng, từ đó giải quyết hiệu quả vấn đề dữ liệu thưa. Việc tính toán mức độ tương tự giữa các cặp sản phẩm dựa vào mô hình đồ thị đề xuất được thực hiện tương tự; 2) Đưa ra đánh giá dự đoán của người dùng với sản phẩm mới trong điều kiện ngữ cảnh cụ thể theo phương pháp k-láng giềng gần nhất, đây là phương pháp được đánh giá là đơn giản trong cài đặt để có thể triển khai thực tế. 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.

Trong thời gian tới, tác giả 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.

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/app712111.
- [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, "Hortling 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.

A CONTEXTUAL MODELING METHOD FOR CONTEXT-AWARE RECOMMENDER SYSTEMS

Abstract: Recommender systems are considered as an active filtering systems acting like decision support systems. The systems has a main purpose as providing users with recommendations on information, products and services that best match their individual preferences. Among approaches in building recommender systems, the context-based approach has been proven to improve quality of suggestions significantly in many practical applications compared to the traditional approaches. However, 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 contextual modeling method that allow fully integrated context situations and resolve effectively the data sparseness problem. In there, relationships between users, items and contexts are represented based on an unique graph model. Then the sparse data problem is solved by fully exploiting the direct and indirect relationships between data objects 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.

Keyword: Context-aware recommender system - CARS; Context-aware collaborative filtering - CACF; Context; Contextual modeling.



Đỗ Thị Liên, Nhận bằng tốt nghiệp đại học, thạc sỹ 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, phát triển ứng dụng đa phương tiện.