

# MỘT PHƯƠNG PHÁP KÍCH HOẠT LAN TRUYỀN TRÊN ĐỒ THỊ CHO HỆ TƯ VẤN CỘNG TÁC THEO NGỮ CẢNH

Đỗ Thị Liên

Học viện Công nghệ Bưu chính Viễn thông

**Tóm tắt:** Hệ thống tư vấn theo ngữ cảnh được thiết kế đặc biệt để khai thác thông tin ngữ cảnh trong quá trình đề xuất sản phẩm phù hợp với người dùng, nhằm đạt hiệu quả cao hơn so với các hệ thống tư vấn truyền thống không sử dụng ngữ cảnh. Một thách thức lớn trong xây dựng hệ tư vấn là dữ liệu thưa, ảnh hưởng đến chất lượng đề xuất. Mặc dù việc sử dụng ngữ cảnh giúp cung cấp thông tin hữu ích nhưng cũng làm tăng độ thưa thớt dữ liệu và độ phức tạp tính toán. Trong nghiên cứu này, tác giả đề xuất một phương pháp mới cho hệ tư vấn theo ngữ cảnh nhằm tích hợp toàn diện thông tin ngữ cảnh, giảm thiểu tác động của vấn đề dữ liệu thưa thớt và duy trì độ phức tạp tính toán thấp. Điều này được giải quyết thông qua việc kết hợp kỹ thuật phân chia người dùng theo ngữ cảnh và phương pháp kích hoạt lan truyền trên đồ thị. Các thử nghiệm thực tế trên một số tập dữ liệu cho thấy phương pháp đề xuất nâng cao đáng kể độ chính xác dự đoán và khả thi về thời gian để triển khai thực tế khi so sánh với các phương pháp tư vấn cộng tác theo ngữ cảnh cơ sở trước đó.

**Từ khóa:** Hệ tư vấn theo ngữ cảnh (Context-Aware Recommender Systems - CARS); Kích hoạt lan truyền (Spreading Activation); Lọc cộng tác theo ngữ cảnh (Context-Aware Collaborative Filtering - CACF); Phân tách người dùng theo ngữ cảnh (User Splitting); Nhận thức ngữ cảnh (Context Awareness)

## I. MỞ ĐẦU

Hệ tư vấn là một hệ thống lọc chủ động nhằm cá nhân hóa các đề xuất sản phẩm, thông tin hoặc dịch vụ phù hợp với sở thích cá nhân của người dùng dựa trên dữ liệu tương tác trước đó của họ. Đứng trước vấn đề quá tải thông tin trên Internet hiện nay, các hệ thống như vậy thực sự đã và đang mang lại nhiều lợi ích to lớn điển hình như Youtube, Amazon, Twitter, Facebook, Lazada... Các mô hình học từ dữ liệu trong hệ tư vấn truyền thống thường được xây dựng dựa trên ba phương pháp chính, đó là phương pháp lọc cộng tác, phương pháp lọc theo nội dung và phương pháp lọc kết hợp [1]. Trong đó, các phương pháp này tập trung vào việc mô hình hóa thông tin người dùng, sản phẩm và đánh giá của người dùng với sản phẩm thông qua ma trận đánh giá, từ đó sinh ra gợi ý sản phẩm mới phù hợp với người dùng. Trong các phương pháp này, lọc cộng tác được xem là dễ dàng để triển khai và có khả năng xử lý mọi loại dữ liệu, đặc biệt là dữ liệu đa phương tiện [1]. Các nghiên cứu đã

có về lọc cộng tác ban đầu thường chỉ chú trọng vào hai yếu tố chính là người dùng và sản phẩm để tiến hành huấn luyện và đưa ra gợi ý. Bài toán lọc cộng tác khi đó có thể mô hình hóa bằng ma trận đánh giá hai chiều, trong đó các hàng đại diện cho người dùng và các cột đại diện cho sản phẩm, mối quan hệ giữa người dùng và sản phẩm được biểu diễn theo (1) như sau:

$$R_{2D}: Users \times Items \rightarrow Ratings \quad (1)$$

Tuy nhiên, sở thích người dùng thay đổi theo ngữ cảnh nên xu hướng trong các hệ tư vấn hiện nay là tích hợp thêm các yếu tố ngữ cảnh này để cung cấp đề xuất chính xác hơn [2], [3], [4]. Theo hướng tiếp cận lọc cộng tác, bài toán tư vấn xử lý theo hướng tiếp cận lọc cộng tác theo ngữ cảnh khi đó được biểu diễn dựa trên ma trận đánh giá đa chiều theo (2) [5] như sau:

$$R_{MD}: Users \times Items \times Contexts \rightarrow Ratings \quad (2)$$

Giả sử hệ thống thu thập được các thông tin về N người dùng, được ký hiệu là  $U = \{u_1, u_2, \dots, u_N\}$  và M sản phẩm, ký hiệu là  $P = \{p_1, p_2, \dots, p_M\}$  và K chiều ngữ cảnh  $C = \{C_1, C_2, \dots, C_K\}$ , mỗi chiều ngữ cảnh có tương ứng  $N_{c_1}, N_{c_2}, \dots, N_{c_K}$  điều kiện ngữ cảnh. Mục tiêu của hệ thống lọc cộng tác theo ngữ cảnh là dự đoán đánh giá và đề xuất các sản phẩm mới cho người dùng trong các ngữ cảnh cụ thể. Để dễ hình dung, Bảng 1 minh họa một ví dụ về ma trận đánh giá đa chiều trong hệ thống tư vấn cộng tác theo ngữ cảnh, bao gồm 2 người dùng  $U = \{u_1, u_2\}$ , 3 sản phẩm  $P = \{p_1, p_2, p_3\}$ , cùng với thông tin về các chiều ngữ cảnh như sau:

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

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

Các nghiên cứu về hệ tư vấn theo ngữ cảnh chỉ ra có hướng tiếp cận chính: lọc trước ngữ cảnh, lọc sau ngữ cảnh và mô hình hóa ngữ cảnh [2], [3]. Thực tế không có phương pháp nào hoàn hảo cho mọi trường hợp dữ liệu, tuy nhiên phương pháp lọc trước ngữ cảnh và mô hình hóa ngữ cảnh đã thu hút được sự chú ý lớn từ giới nghiên cứu, với nhiều công bố chứng minh hiệu quả cao trong nhiều tình huống khác nhau. Mặc dù có hiệu quả cao, các phương pháp này

Tác giả liên hệ: Đỗ Thị Liên,

Email: [liendt@ptit.edu.vn](mailto:liendt@ptit.edu.vn)

Đến tòa soạn: 8/2024, chỉnh sửa: 9/2024, chấp nhận đăng: 9/2024.

vấn gặp thách thức về tính thừa thớt dữ liệu và độ phức tạp tính toán [3], [6]. Trong bài báo này, tác giả đưa ra một đề xuất mới thuộc hướng tiếp cận lọc trước ngữ cảnh nhằm giải quyết các vấn đề nêu trên bằng cách kết hợp kỹ thuật phân tách người dùng theo ngữ cảnh cải tiến và phương pháp kích hoạt lan truyền trên đồ thị. Các thử nghiệm cho thấy hiệu quả của phương pháp đề xuất trong việc tích hợp đầy đủ thông tin ngữ cảnh và giải quyết hiệu quả vấn đề dữ liệu thừa từ đó cải thiện độ chính xác dự đoán khi so sánh với các phương pháp tư vấn cộng tác theo ngữ cảnh cơ sở. Ngoài ra, phương pháp đề xuất được đánh giá là có độ phức tạp thời gian và không gian tương đối hiệu quả và phù hợp cho các hệ thống tư vấn trong nhiều trường hợp.

Đề tập trung vào phương pháp đề xuất, phần II sẽ giới thiệu các nghiên cứu liên quan. Sau đó, phương pháp đề xuất sẽ được mô tả trong phần III. Phần IV sẽ trình bày chi tiết phương pháp thực nghiệm và đánh giá. Cuối cùng, phần V sẽ đưa ra kết luận và hướng phát triển trong tương lai.

## II. CÁC NGHIÊN CỨU LIÊN QUAN

Trong những năm gần đây, các hệ thống gợi ý dựa trên ngữ cảnh đã nhận được sự quan tâm đáng kể trong cộng đồng nghiên cứu. Những hệ thống này không chỉ dựa vào thông tin người dùng và sản phẩm mà còn tích hợp thêm thông tin ngữ cảnh như thời gian, địa điểm, và tình huống xung quanh. Việc tích hợp này giúp cải thiện hiệu suất và tính chính xác của các gợi ý bằng cách hiểu rõ hơn về tình huống cụ thể của người dùng khi sử dụng sản phẩm.

Việc khai thác thông tin ngữ cảnh trong hệ thống tư vấn được giới thiệu đầu tiên trong nghiên cứu của Adomavicius và Tuzhilin [7] và từ đó, đã có nhiều nghiên cứu khác nhau được thực hiện nhằm giải quyết các vấn đề liên quan đến hệ thống gợi ý theo ngữ cảnh. Các loại ngữ cảnh khác nhau đã được sử dụng trong các hệ thống gợi ý, tùy thuộc vào ứng dụng cụ thể và sự sẵn có của dữ liệu chứa ngữ cảnh. Các yếu tố ngữ cảnh phổ biến bao gồm thời gian [8], [9], địa điểm [10], [11], thứ tự thời gian của các mục trong các phiên ngắn [12], [13]. Hệ thống gợi ý theo ngữ cảnh (CARS) đã chứng minh được hiệu quả của mình trong nhiều lĩnh vực khác nhau, bao gồm mạng xã hội [14], gợi ý địa điểm quan tâm [15], [16], thương mại điện tử [17], [18], gợi ý dịch vụ web [19]...

Có ba hướng tiếp cận chính cho hệ thống gợi ý theo ngữ cảnh: lọc trước ngữ cảnh (Contextual Prefiltering), lọc sau ngữ cảnh (Contextual Postfiltering) và mô hình hóa ngữ cảnh (Contextual Modeling) [4], [6], [7].

- **Lọc trước theo ngữ cảnh:** Sử dụng thông tin ngữ cảnh để lọc dữ liệu ban đầu, chỉ giữ lại những dữ liệu phù hợp với ngữ cảnh yêu cầu. "Phân tách người dùng theo ngữ cảnh" là một phương pháp điển hình theo hướng tiếp cận này, cho phép phân chia người dùng hiện thời thành các người dùng giả lập dựa trên một điều kiện ngữ cảnh có độ lợi thông tin cao nhất đối với các đánh giá. Một số phương pháp lọc trước ngữ cảnh đã được đề xuất bao gồm: phương pháp phân chia dữ liệu (Splitting), phương pháp sử dụng một phần nhỏ của các chiều ngữ cảnh (Context Relaxation) và phương pháp lọc dựa trên ngữ nghĩa (Semantic Filtering). Dữ liệu sau khi được lọc sẽ được sử dụng để huấn luyện và đưa ra gợi ý. Trong quá trình này, các phương pháp lọc cộng tác truyền thống có thể được áp dụng trực tiếp cho hệ thống gợi ý dựa trên ngữ cảnh. Các phương pháp như UserKNN, ItemKNN, Matrix Factorization, và SLIM sẽ được sử

dụng sau bước lọc trước ngữ cảnh để tạo ra các đề xuất sản phẩm cho người dùng trong các tình huống ngữ cảnh khác nhau.

- **Lọc sau theo ngữ cảnh:** Sử dụng toàn bộ ma trận đánh giá mà không loại bỏ các chiều ngữ cảnh trong quá trình huấn luyện và gợi ý, sau đó lọc lại các kết quả dựa trên ngữ cảnh hiện tại. Do đó, các phương pháp tư vấn cộng tác theo ngữ cảnh thuộc hướng lọc sau ngữ cảnh cũng có thể sử dụng các kỹ thuật tư vấn cộng tác truyền thống, tương tự như phương pháp lọc trước ngữ cảnh. Trong trường hợp này, thông tin ngữ cảnh không được kết hợp vào quá trình huấn luyện.
- **Mô hình hóa ngữ cảnh:** Thông tin ngữ cảnh, người dùng, và sản phẩm được biểu diễn trong cùng một mô hình để thực hiện huấn luyện và gợi ý. Theo đó, phương pháp tiếp cận mô hình hóa ngữ cảnh trực tiếp kết hợp thông tin ngữ cảnh, người dùng và sản phẩm trong cùng một mô hình, sử dụng ma trận đánh giá đa chiều cho quá trình huấn luyện và gợi ý. Các phương pháp mô hình hóa ngữ cảnh được chia thành hai nhóm: mô hình hóa ngữ cảnh độc lập và mô hình hóa ngữ cảnh phụ thuộc. Tensor Decomposition [20] là một ví dụ tiêu biểu của mô hình hóa ngữ cảnh độc lập, cho phép biểu diễn các thành phần trong một không gian đa chiều, tuy nhiên, nó gặp khó khăn với việc xử lý dữ liệu khi số lượng chiều ngữ cảnh tăng cao. Ngược lại, mô hình hóa ngữ cảnh phụ thuộc mô tả và khai thác mối quan hệ giữa người dùng, sản phẩm và ngữ cảnh, mang lại kết quả tốt hơn nhưng cũng phức tạp hơn về mặt tính toán và mở rộng. Một số nghiên cứu gần đây theo hướng tiếp cận mô hình hóa ngữ cảnh là tập trung vào khai thác mô hình học sâu trong xây dựng hệ tư vấn, điển hình như nghiên cứu [21] kết hợp mạng nơ-ron và bộ mã hóa tự động autoencoder để trích xuất đặc trưng và đưa ra dự đoán đánh giá từ dữ liệu đầu vào, hay nghiên cứu [22] đề xuất sử dụng kỹ thuật phân rã ma trận (Matrix Factorization) cùng với ma trận chuyển tiếp sở thích để nắm bắt và phản ánh những thay đổi trong sở thích của người dùng theo thời gian. Mặc dù các nghiên cứu này cũng chỉ ra rằng việc kết hợp các phương pháp lọc cộng tác với các kỹ thuật học sâu có thể nâng cao đáng kể hiệu suất của các hệ thống gợi ý nhưng một vài khó khăn thách thức điển hình đối với các hệ tư vấn khi áp dụng học sâu là yêu cầu về dữ liệu đầu vào lớn, chi phí tính toán cao, khả năng giải thích kém cũng như khó khăn trong việc tinh chỉnh mô hình. Thực nghiệm cho thấy các phương pháp theo hướng tiếp cận mô hình hóa ngữ cảnh thường có độ phức tạp lớn và khả năng mở rộng khó khăn hơn các phương pháp lọc trước và sau ngữ cảnh, đặc biệt khi số chiều ngữ cảnh tăng lên.

Trong bài báo này, tác giả đề xuất một phương pháp mới cho hệ thống tư vấn theo ngữ cảnh thuộc hướng tiếp cận lọc trước ngữ cảnh. Phương pháp đề xuất được thực hiện trên cơ sở phương pháp kích hoạt lan truyền trên đồ thị PageRank [23]. Thuật toán PageRank là một thuật toán quan trọng trong tìm kiếm web để xác định mức độ quan trọng của các trang dựa trên cấu trúc liên kết của chúng, là một thuật toán cơ sở quan trọng được áp dụng thành công trong các máy tìm kiếm của Google. Cơ chế thực hiện của thuật toán này thực hiện việc lan truyền và cập nhật xếp hạng cho các đỉnh trong đồ thị từ đó đánh giá và phân loại mức độ quan trọng cho các đỉnh đó. PageRank ban đầu được phát triển để xếp hạng các trang web trên Internet, sau

này đã được mở rộng và áp dụng trong nhiều lĩnh vực khác nhau, bao gồm cả hệ tư vấn [24]. Việc triển khai thuật toán PageRank giúp hệ thống tư vấn cải thiện độ chính xác và cá nhân hóa bằng cách xem xét không chỉ các đánh giá trực tiếp mà còn cả mối quan hệ gián tiếp giữa các đối tượng dựa trên mô hình đồ thị. Trong bối cảnh này, một số nghiên cứu đã được đưa ra. Điển hình như một số nghiên cứu sau: Nghiên cứu [25] đề xuất một phương pháp mới gọi là "UserRank" sử dụng PageRank để xếp hạng người dùng dựa trên mối quan hệ giữa họ và các sản phẩm mà họ đánh giá. Bằng cách tích hợp UserRank vào việc tính toán độ tương đồng giữa các sản phẩm, phương pháp này cải thiện đáng kể độ chính xác và độ ổn định của các gợi ý so với các thuật toán lọc cộng tác truyền thống dựa trên độ đo Cosine; Nghiên cứu [26][27] cải thiện khả năng gợi ý thông qua việc tính toán mức độ liên quan giữa các đối tượng và ưu tiên các đối tượng có liên kết cao trong hệ thống gợi ý dựa trên dữ liệu liên kết, tuy nhiên tập trung cho các hệ tư vấn truyền thống chưa xem xét đầy đủ các yếu tố ngữ cảnh; Trong nghiên cứu [28] thuật toán PageRank được điều chỉnh để cải thiện quá trình lọc cộng tác (CF) trong các thư viện số; Nghiên cứu [29] đưa ra phương pháp sử dụng Personalized PageRank trong hệ thống gợi ý dựa trên ngữ cảnh mặc dù mang lại hiệu quả tuy nhiên vẫn gặp hạn chế trong việc xử lý các tình huống ngữ cảnh phức tạp và phụ thuộc vào cấu trúc đồ thị, yêu cầu tính toán phức tạp và khả năng gây bias về các sản phẩm phổ biến. Nhìn chung các nghiên cứu đã có này chủ yếu tập trung vào việc áp dụng thuật toán PageRank vào hệ tư vấn không sử dụng ngữ cảnh, mặc dù nghiên cứu [29] đã chú trọng sử dụng PageRank trong hệ thống gợi ý dựa trên ngữ cảnh tuy nhiên nghiên cứu này tiếp cận hướng mô hình hóa ngữ cảnh còn gặp phải độ phức tạp tính toán cao và khó khăn trong việc tích hợp các tình huống ngữ cảnh phức tạp.

Để giảm thiểu những hạn chế nêu trên, tác giả đưa ra đề xuất một phương pháp lọc cộng tác theo ngữ cảnh mới cho hệ tư vấn, đây là phương pháp được tiếp cận theo hướng lọc trước ngữ cảnh. Phương pháp này tích hợp ngữ cảnh thông qua một thủ tục phân tách người dùng theo ngữ cảnh cải tiến và giải quyết vấn đề dữ liệu thưa thớt bằng phương pháp kích hoạt lan truyền trên đồ thị. Trong đó thủ tục phân tách người dùng theo ngữ cảnh đề xuất sử dụng tất cả các chiều ngữ cảnh để tạo ra các người dùng giả lập, giúp tích hợp đầy đủ thông tin ngữ cảnh vào quá trình gợi ý. Tuy nhiên, việc này cũng làm gia tăng tính thưa thớt của dữ liệu. Để khắc phục điều đó, tác giả sử dụng mô hình đồ thị để biểu diễn tất cả các mối quan hệ trực tiếp và gián tiếp giữa người dùng và sản phẩm. Từ biểu diễn đồ thị này, tác giả áp dụng phương pháp kích hoạt lan truyền trên đồ thị PageRank [23] vào hệ tư vấn để cập nhật xếp hạng mới của người dùng đối với sản phẩm trong hệ thống, từ đó đề xuất những sản phẩm có xếp hạng cao cho họ. Việc đề xuất phương pháp thuộc hướng tiếp cận lọc trước ngữ cảnh cũng được đánh giá là đơn giản để cài đặt, giải pháp đưa ra được đánh giá là có độ phức tạp tính toán thấp để dễ dàng triển khai trên thực tế. Nội dung cụ thể của giải pháp đề xuất được trình bày chi tiết trong Mục III.

### III. PHƯƠNG PHÁP ĐỀ XUẤT

Phương pháp đề xuất được triển khai bằng cách kết hợp hai kỹ thuật chính: 1) Phân chia người dùng theo ngữ cảnh cải tiến và 2) Lọc cộng tác sử dụng phương pháp kích hoạt lan truyền trên đồ thị.

#### 3.1. Phân tách người dùng theo ngữ cảnh cải tiến

Phương pháp phân tách người dùng theo ngữ cảnh đề xuất cải tiến từ phương pháp phân tách người dùng theo ngữ cảnh nguyên thủy User Splitting trước đây [30] nhằm khai thác thông tin ngữ cảnh hiệu quả hơn. Trong đề xuất của mình, thay vì chỉ sử dụng một chiều ngữ cảnh duy nhất như trước đây, tác giả sẽ kết hợp tất cả các chiều ngữ cảnh tham gia vào quá trình lọc trước theo ngữ cảnh. Các bước thực hiện như sau:

##### • Bước 1. Tạo tập người dùng giả lập

Tập người dùng giả lập  $Y$  được tạo ra bằng cách lấy tích Đề-các của tập người dùng  $U$  với tất cả  $K$  chiều ngữ cảnh  $C_1, C_2, \dots, C_K$ . Cụ thể:  $Y = U \times C_1 \times C_2 \times \dots \times C_K$ . Số các điều kiện ngữ cảnh trong tập  $C_i$  là  $N_{C_i}$  với  $i=1 \dots K$ , số lượng người dùng trong tập  $Y$  là  $H$  với  $H = N \times N_{C_1} \times N_{C_2} \times \dots \times N_{C_K}$ .

##### • Bước 2. Chuyển đổi ma trận đánh giá

Chuyển đổi ma trận đánh giá đa chiều thành ma trận đánh giá hai chiều bằng cách loại bỏ tập ngữ cảnh và thay thế tập người dùng ban đầu bằng tập người dùng giả lập. Như vậy, phương pháp phân tách người dùng theo ngữ cảnh cải tiến này giúp tích hợp đầy đủ thông tin ngữ cảnh vào hệ thống mà không làm thay đổi các điều kiện của bài toán.

Ví dụ: Việc áp dụng quy trình phân tách người dùng theo ngữ cảnh cải tiến vào ma trận đánh giá đa chiều của lọc cộng tác theo ngữ cảnh (Bảng 1) giúp tạo ra một ma trận đánh giá hai chiều (Bảng 2). Cụ thể với 2 người dùng và 12 kịch bản ngữ cảnh đưa ra trong Bảng 1, phương pháp phân tách người dùng theo ngữ cảnh cải tiến tạo ra tổng cộng 24 người dùng giả lập ( $y_1, y_2, \dots, y_{24}$ ) ở Bảng 2. Chẳng hạn như  $y_1$ , được tạo ra thông qua sự kết hợp của người dùng  $u_1$  và các điều kiện ngữ cảnh cụ thể ("Cuối tuần", "Ở nhà", "Trẻ em"). Ma trận đánh giá hai chiều sau đó trình bày các cặp người dùng-sản phẩm với các đánh giá được hiển thị ở các hàng trên của ma trận, còn các cặp không có đánh giá được ẩn đi để tiết kiệm không gian.

Thực tế là các đánh giá ban đầu của người dùng đối với sản phẩm trong các tình huống ngữ cảnh thường khá hạn chế về số lượng, dẫn đến mức độ thưa thớt cao trong ma trận đánh giá. Khi áp dụng quy trình phân tách người dùng theo ngữ cảnh cải tiến vào ma trận đa chiều, việc tạo ra người dùng giả lập càng làm tăng thêm sự thưa thớt này trong ma trận đánh giá hai chiều. Kết quả là, ma trận trở nên thưa thớt hơn, với nhiều cặp người dùng-sản phẩm không có đánh giá. Sự gia tăng dữ liệu thừa tạo ra thách thức trong việc tạo ra các đề xuất chính xác. Để giải quyết vấn đề này, phương pháp kích hoạt lan truyền trên đồ thị PageRank [23] được đề xuất trong bài báo này nhằm cập nhật xếp hạng mới cho người dùng và đưa ra đề xuất các sản phẩm chưa được đánh giá nhưng có xếp hạng cao sau quá trình học cho họ.

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

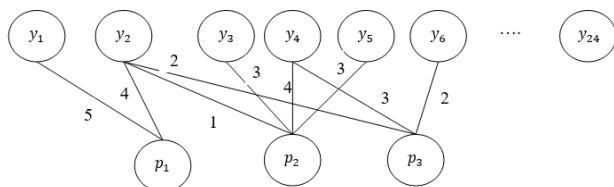
| Y     | P     | R |
|-------|-------|---|
| $y_1$ | $p_1$ | 5 |
| $y_2$ | $p_1$ | 4 |
| $y_2$ | $p_3$ | 2 |
| $y_3$ | $p_2$ | 3 |
| $y_4$ | $p_2$ | 4 |
| $y_5$ | $p_2$ | 3 |
| $y_6$ | $p_3$ | 2 |
| $y_2$ | $p_2$ | 1 |

|          |       |     |
|----------|-------|-----|
| $y_4$    | $p_3$ | 3   |
| ...      | ...   | ... |
| $y_{24}$ | $p_3$ | 0   |

### 3.2. Loại cộng tác sử dụng phương pháp kích hoạt lan truyền trên đồ thị

#### 3.2.1. Biểu diễn đồ thị cho loại cộng tác

Từ ma trận đánh giá hai chiều thu được trong Bảng 2, tác giả biểu diễn mối quan hệ giữa tập người dùng giả lập và tập sản phẩm dựa trên đồ thị  $G=\langle V, E \rangle$ . Trong đó, tập đỉnh  $V$  gồm tập người dùng giả lập và tập sản phẩm ( $V=Y \cup P$ ) và tập cạnh  $E$  thể hiện liên kết giữa người dùng giả lập và sản phẩm theo Bảng 2, với mỗi cạnh  $e \in E$  có dạng  $e = (i, x)$ , với  $i \in Y$  và  $x \in P$ . Trọng số mỗi cạnh trên đồ thị được thiết lập chính là đánh giá của người dùng với sản phẩm tương ứng từ ma trận đánh giá. Biểu diễn đồ thị cho loại cộng tác được minh họa trong Hình 1.



Hình 1. Biểu diễn đồ thị cho loại cộng tác

#### 3.2.2. Dự đoán đánh giá của người dùng giả lập với sản phẩm dựa trên biểu diễn đồ thị

Mô hình đồ thị biểu diễn cho loại cộng tác ở trên cho phép thể hiện đầy đủ các mối quan hệ trực tiếp và gián tiếp giữa người dùng giả lập và sản phẩm. Do đó, việc xây dựng mô hình học dữ liệu dựa trên biểu diễn đồ thị này giúp giải quyết hiệu quả vấn đề dữ liệu thưa. Trong nghiên cứu này, tác giả ứng dụng phương pháp kích hoạt lan truyền trên đồ thị PageRank [23] vào hệ tư vấn với mục tiêu là xếp hạng các sản phẩm dựa trên mức độ liên quan và sở thích của người dùng.

Theo nghiên cứu gốc về thuật toán PageRank [23] xếp hạng PageRank của mỗi trang web  $p_i$  được xác định dựa trên giá trị PageRank của các trang liên kết đến nó, thực hiện theo công thức sau:

$$rank(p_i) = \frac{1-d}{N} + d \sum_{p_j \in L(p_i)} \frac{rank(p_j)}{L(p_j)} \quad (3)$$

#### Đầu vào:

- Đồ thị mô tả mối quan hệ giữa người dùng và sản phẩm được biểu diễn dưới dạng ma trận đánh giá 2 chiều giữa người dùng giả lập và sản phẩm  $R^{(0)} = \{r_{ix}\}$
- $U$ : Tập hợp người dùng
- $P$ : Tập hợp sản phẩm
- $d$ : Hệ số làm giảm
- $max\_iteration$ : Số lần lặp tối đa
- $\epsilon$ : Ngưỡng hội tụ

#### Đầu ra :

- Rank: Bảng xếp hạng các sản phẩm theo từng người dùng giả lập

#### Các bước tiến hành:

##### 1. Khởi tạo:

```
For each user u in U:
    For each product p in P:
```

Trong đó:

- $PR_{p_i}$ : Là giá trị xếp hạng PageRank của trang web  $p_i$ , cho biết tầm quan trọng của trang  $p_i$  trong hệ thống.
- $d$ : Hệ số làm giảm (Thường là 0.85), thể hiện xác suất mà PageRank tiếp tục di chuyển từ trang web này đến trang web khác thông qua các liên kết.
- $N$ : Tổng số trang trong mạng lưới.
- $p_j$ : Các trang liên kết tới trang  $p_i$
- $L(p_j)$ : Số lượng liên kết đi ra từ trang  $p_j$
- $d \sum_{p_j \in L(p_i)} \frac{PR(p_j)}{L(p_j)}$ : Tính toán xếp hạng của  $p_i$  dựa trên xếp hạng của các trang web liên kết với nó.
- $\frac{1-d}{N}$ : Phần này đảm bảo rằng tất cả các trang web đều có một mức độ quan trọng nhất định, ngay cả khi chúng không có liên kết nào dẫn đến. Đây là giá trị nhỏ xếp hạng ngẫu nhiên gán cho mỗi trang web nhằm đảm bảo rằng tất cả các trang web đều có cơ hội được xếp hạng.

Trong bối cảnh của hệ tư vấn, để xếp hạng các sản phẩm dựa trên mức độ liên quan và sở thích của người dùng, giá trị PageRank cần cá nhân hóa theo từng người dùng, từ đó đưa ra các đề xuất phù hợp hơn. Như vậy công thức PageRank trong hệ tư vấn được điều chỉnh như sau:

$$rank_u(p_i) = \frac{1-d}{N} + d \sum_{p_j \in L(p_i)} \frac{rank_u(p_j)}{L(p_j)} \quad (4)$$

Trong đó:  $rank_u(p_i)$  là giá trị xếp hạng PageRank của sản phẩm  $p_i$  đối với người dùng  $u$ .

Dựa trên công thức (4) trên, tác giả tiến hành đề xuất thuật toán dự đoán đánh giá của người dùng đối với sản phẩm bằng cách lan truyền và điều chỉnh xếp hạng sản phẩm qua nhiều bước. Ứng dụng thuật toán PageRank cho phép tận dụng thông tin từ toàn bộ đồ thị người dùng – sản phẩm để giúp xác định các sản phẩm có xếp hạng toàn cục cao nhất nhằm đưa ra gợi ý cho người dùng, nhờ đó nâng cao chất lượng và độ chính xác của các đề xuất. Thuật toán được miêu tả chi tiết dưới đây.

```
# Khởi tạo giá trị PageRank ban đầu bằng nhau cho tất cả các sản phẩm
Rank[u][p] = 1/|P|

2. Lan truyền nhân xếp hạng PageRank cho từng sản phẩm
For t from 1 to max_iteration:
    For each user u in U:
        For each product p in P:
            inbound_sum = 0
            For each product q in inbound_links(p):
                #Tất cả các sản phẩm liên kết đến p
                inbound_sum += previous_rank[u][q] / out_degree(q)
            Rank[u][p] = (1 - d) / |P| + d * inbound_sum

# Kiểm tra điều kiện hội tụ
delta = 0
```

```

For each user u in U:
    For each product p in P:
        delta += abs(Rank[u][p] -
previous_rank[u][p])
    If delta < epsilon:
        Break # Hội tụ

```

**Thuật toán 1. Thuật toán PageRank dự đoán đánh giá của người dùng giả lập với sản phẩm**

### 3.2.3. Sinh tư vấn

Thuật toán PageRank dự đoán đánh giá của người dùng giả lập cho sản phẩm với đầu ra là bảng xếp hạng các sản phẩm theo từng người dùng giả lập *Rank*, đây là cơ sở để sinh tư vấn sản phẩm cá nhân hóa cho người dùng trong ngữ cảnh.

**Đầu vào**

- Ma trận đánh giá đa chiều (chứa thông tin ngữ cảnh)
- $u_a \in U$  là người dùng hiện thời cần được tư vấn
- $c_t \in (C_1 \times C_2 \times \dots \times C_K)$  là tình huống ngữ cảnh ứng với  $u_a$
- $K$  là số lượng sản phẩm cần tư vấn cho  $u_a$

**Đầu ra:**

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

**Các bước tiến hành:**

1. Phân tách người dùng theo ngữ cảnh cải tiến (Mục 3.1)
2. Lọc cộng tác sử dụng phương pháp kích hoạt lan truyền trên đồ thị PageRank (Theo thuật toán 1)
3. Sinh tư vấn cho người dùng hiện thời  $u_a$  trong ngữ cảnh  $c_t$ 
  - Chuyển đổi bảng Rank thể hiện xếp hạng giữa người dùng giả lập với sản phẩm về ma trận đánh giá đa chiều
  - Sắp xếp thứ tự xếp hạng của  $u_a$  với  $c_t$  theo thứ tự giảm dần
  - Chọn  $K$  sản phẩm có xếp hạng cao nhất tư vấn người dùng hiện thời  $u_a$  trong tình huống ngữ cảnh  $c_t$ .

**Thuật toán 2. Thuật toán US\_ActiveGraph**

Đánh giá độ phức tạp về mặt thời gian của thuật toán đề xuất phụ thuộc trực tiếp vào độ phức tạp của thuật toán 1. Căn cứ theo nguyên lý cộng và nguyên lý nhân cho đánh giá độ phức tạp của các cấu trúc lệnh, ta xác định được thuật toán này có độ phức tạp là:  $O(T \times H \times M \times |L(p)|)$ . Trong đó  $T$  là số lần lặp,  $H$  là số lượng người dùng giả lập,  $M$  là số lượng sản phẩm và  $|L(p)|$  là số lượng sản phẩm liên kết tới mỗi sản phẩm. Từ cơ sở này, thời gian tính toán của phương pháp đề xuất được đánh giá có độ phức tạp thấp hơn so với các phương pháp khác như mô hình hóa ngữ cảnh hoặc các thuật toán học sâu, đặc biệt hiệu quả trong những trường hợp với dữ liệu ngữ cảnh phức tạp và khả năng xử lý tốt vấn đề dữ liệu thưa thớt. Đồng thời thuật toán cũng được đánh giá là không khó khăn trong việc tiếp cận và đơn giản trong cài đặt để triển khai trên thực tế.

## IV. THỰC NGHIỆM VÀ KẾT QUẢ

### 4.1. Dữ liệu thực nghiệm

Trong nghiên cứu này, tác giả sử dụng hai bộ dữ liệu thực nghiệm phổ biến là InCarMusic và DepaulMovie để đánh giá hiệu quả của phương pháp đề xuất.

- Bộ dữ liệu DepaulMovie [31] chứa 5043 đánh giá từ 97 người dùng cho 79 bộ phim. Bộ dữ liệu này được đặc trưng bởi ba chiều ngữ cảnh: *Time*, *Location*, *Companion*. Mức độ thưa thớt của các đánh giá trong bộ dữ liệu này là 94.516%.
- Bộ dữ liệu InCarMusic [31] bao gồm 3938 đánh giá từ 1042 người dùng về 139 album. Bộ dữ liệu này được đặc trưng bởi tám chiều ngữ cảnh: *Driving style*, *Road type*, *Landscape*, *Sleepiness*, *Traffic conditions*, *Mood*, *Weather*, *Natural Phenomena*. Mức độ thưa thớt của các đánh giá trong bộ dữ liệu này là 99.99%.

### 4.2. Cài đặt thực nghiệm

**Độ đo thực nghiệm:** Trong nghiên cứu này, bài báo tập trung vào đề xuất top-N với mục đích giới thiệu  $N$  sản phẩm có xếp hạng cao nhất cho người dùng. Để so sánh phương pháp này với các thuật toán đề xuất top-N khác, chúng tôi sử dụng các độ đo Precision@N ( $P@N$ ) và Mean Average Precision (MAP@N) là những độ đo được rộng rãi được sử dụng trong việc đề xuất top-N. Tác giả sử dụng  $N=10$  trong tất cả các thí nghiệm.

**Tham số thuật toán:** Hệ số làm giảm  $d$  trong thuật toán PageRank được thiết lập là 0.85. Số lần lặp tối đa của thuật toán được thiết lập là 1000, và ngưỡng hội tụ  $\epsilon$  được thiết lập là 0.0001. Việc lựa chọn giá trị cho các tham số này trong thuật toán PageRank dựa trên sự kết hợp giữa lý thuyết, kinh nghiệm thực tế và các nghiên cứu trước đây. Những giá trị này không chỉ đảm bảo tính ổn định và độ chính xác của kết quả mà còn tối ưu hóa hiệu suất tính toán, phù hợp với đa số các ứng dụng thực tiễn của PageRank [23].

**Các phương pháp tư vấn theo ngữ cảnh được lựa chọn để so sánh:**

- *CAMF\_C* (Context-aware matrix factorization-Context) [32]: Sử dụng phương pháp mô hình hóa ngữ cảnh độc lập dựa trên sự chênh lệch đánh giá giữa các điều kiện ngữ cảnh.
- *CSLIM\_MCS* (Contextual Sparse Linear Method - Multidimensional Context Similarity)[31]: Sử dụng phương pháp tuyến tính thưa SLIM kết hợp với mô hình hóa ngữ cảnh dựa trên độ tương quan đa chiều.
- *UserSplitting-BiasedMF*; *ItemSplitting-BiasedMF*; *UISplitting-BasedMF* [31]: Đây là các phương pháp tư vấn theo ngữ cảnh áp dụng lần lượt phân chia theo người dùng nguyên thủy/ phân chia theo sản phẩm nguyên thủy / kết hợp phân chia theo cả người dùng và sản phẩm nguyên thủy dựa vào ngữ cảnh. Sau đó, hệ thống sử dụng phương pháp Biased Matrix Factorization (BiasedMF) để huấn luyện và gợi ý.
- *UserSplitting\_ActiveGraph*: Kết hợp phân chia người dùng theo ngữ cảnh nguyên thủy (sử dụng 1 chiều ngữ cảnh) và phương pháp kích hoạt lan truyền trên đồ thị PageRank.
- *US\_ActiveGraph*: Phương pháp đề xuất bởi bài báo.

### 4.3. Kết quả thực nghiệm

Bảng III. Kết quả thực nghiệm trên tập dữ liệu DepaulMovie

| Methods                   | Precision@10 | MAP@10       |
|---------------------------|--------------|--------------|
| CAMF_C                    | 0.088        | 0.147        |
| CSLIM_MCS                 | 0.085        | 0.159        |
| UserSplitting-BiasedMF    | 0.089        | 0.166        |
| ItemSplitting-BiasedMF    | 0.086        | 0.16         |
| UISplitting-BiasedMF      | 0.084        | 0.156        |
| UserSplitting_ActiveGraph | 0.092        | 0.162        |
| <b>US_ActiveGraph</b>     | <b>0.103</b> | <b>0.191</b> |

Bảng IV. Kết quả thực nghiệm trên tập dữ liệu InCarMusic

| Methods                   | Precision@10 | MAP@10       |
|---------------------------|--------------|--------------|
| CAMF_C                    | 0.032        | 0.127        |
| CSLIM_MCS                 | 0.018        | 0.141        |
| UserSplitting-BiasedMF    | 0.033        | 0.259        |
| ItemSplitting-BiasedMF    | 0.034        | 0.267        |
| UISplitting-BiasedMF      | 0.033        | 0.259        |
| UserSplitting_ActiveGraph | 0.038        | 0.248        |
| <b>US_ActiveGraph</b>     | <b>0.040</b> | <b>0.314</b> |

Dựa trên phân tích kết quả thí nghiệm được trình bày trong Bảng 3 và Bảng 4, có thể rút ra một số nhận xét sau:

- 1) Các phương pháp UserSplitting-BiasedMF, ItemSplitting-BiasedMF, UISplitting-BiasedMF về cơ bản cho lại kết quả Precision@10 và MAP@10 cải thiện tốt hơn so với các phương pháp mô hình hóa theo ngữ cảnh cơ sở. Điều này cho thấy việc tiếp cận lọc trước ngữ cảnh với việc kết hợp phương pháp phân tách theo ngữ cảnh và lọc cộng tác cho lại kết quả tốt hơn phương pháp mô hình hóa ngữ cảnh trong nhiều trường hợp.
- 2) Phương pháp UserSplitting\_ActiveGraph cho thấy giá trị Precision@10 được cải thiện so với các phương pháp tư vấn theo ngữ cảnh cơ sở. Tuy nhiên, kết quả MAP@10 của chúng lại thấp hơn, cho thấy việc chỉ sử dụng một chiều ngữ cảnh trong phương pháp phân tách người dùng kết hợp với phương pháp kích hoạt lan truyền trên đồ thị có thể không phải là giải pháp tối ưu.
- 3) Phương pháp đề xuất US\_ActiveGraph kết hợp phương pháp phân tách người dùng theo ngữ cảnh cải tiến (sử dụng đồng thời nhiều chiều ngữ cảnh) với phương pháp kích hoạt lan truyền trên đồ thị PageRank cho lại Precision@10 và MAP@10 tốt hơn so với các phương pháp tư vấn theo ngữ cảnh cơ sở và UserSplitting\_ActiveGraph sử dụng phân tách người dùng theo một chiều ngữ cảnh. Điều này cho thấy rằng việc tích hợp nhiều chiều ngữ cảnh mang lại thông tin giá trị cho quá trình đề xuất, cải thiện cả tiêu chí Precision@10 và MAP@10. Đồng thời phương pháp đề xuất tỏ ra ít nhạy cảm hơn với dữ liệu thừa thớt so với các phương pháp cơ sở trên cả hai bộ dữ liệu, ngay cả khi tích hợp đầy đủ thông tin ngữ cảnh.

### V. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong nghiên cứu này, tác giả đã đề xuất một phương pháp mới cho hệ thống tư vấn theo ngữ cảnh thuộc hướng

tiếp cận lọc trước theo ngữ cảnh. Phương pháp đề xuất được thực hiện bằng việc kết hợp giữa kỹ thuật phân chia người dùng theo ngữ cảnh và thuật toán lan truyền trên đồ thị. Trong đó, thủ tục phân tách người dùng theo ngữ cảnh cải tiến tạo điều kiện cho việc tích hợp đầy đủ thông tin ngữ cảnh vào quá trình tư vấn, nhưng điều này cũng làm gia tăng tình trạng thừa thớt dữ liệu và độ phức tạp tính toán. Để giải quyết vấn đề này, chúng tôi áp dụng phương pháp kích hoạt lan truyền trên đồ thị PageRank vào hệ tư vấn để thực hiện cập nhật xếp hạng mới cho người dùng đối với sản phẩm trong hệ thống, từ đó đề xuất những sản phẩm có xếp hạng cao cho người dùng. Phương pháp đề xuất được đánh giá là cho phép tích hợp đầy đủ thông tin ngữ cảnh, hạn chế ảnh hưởng của vấn đề dữ liệu thừa, đồng thời thuận lợi cho việc tiếp cận để triển khai trên 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 đó.

Trong thời gian tới, tác giả dự định sẽ tập trung vào tối ưu hóa hiệu suất tính toán để giảm thời gian thực thi, nâng cao khả năng tích hợp ngữ cảnh động, cũng như mở rộng quy mô thực nghiệm trên các bộ dữ liệu lớn hơn.

### TÀI LIỆU THAM KHẢO

- [1] E. Zangerle and C. Bauer, "Evaluating Recommender Systems: Survey and Framework," *ACM Comput Surv*, vol. 55, no. 8, pp. 1–38, Aug. 2023, doi: 10.1145/3556536.
- [2] M. Casillo, F. Colace, D. Conte, M. Lombardi, D. Santaniello, and C. Valentino, "Context-aware recommender systems and cultural heritage: a survey," *J Ambient Intell Humaniz Comput*, vol. 14, no. 4, 2023, doi: 10.1007/s12652-021-03438-9.
- [3] N. M. Villegas, C. Sánchez, J. Díaz-Cely, and G. Tamura, "Characterizing context-aware recommender systems: A systematic literature review," *Knowl Based Syst*, vol. 140, 2018, doi: 10.1016/j.knosys.2017.11.003.
- [4] K. Verbert *et al.*, "Context-aware recommender systems for learning: A survey and future challenges," 2012. doi: 10.1109/TLT.2012.11.
- [5] S. Kulkarni and S. F. Rodd, "Context Aware Recommendation Systems: A review of the state of the art techniques," 2020. doi: 10.1016/j.cosrev.2020.100255.
- [6] X. Meng, Y. Du, Y. Zhang, and X. Han, "A Survey of Context-Aware Recommender Systems: From an Evaluation Perspective," *IEEE Trans Knowl Data Eng*, vol. 35, no. 7, 2023, doi: 10.1109/TKDE.2022.3187434.
- [7] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," *Recommender Systems Handbook, Second Edition*, 2015, doi: 10.1007/978-1-4899-7637-6\_6.
- [8] P. Kefalas and Y. Manolopoulos, "A time-aware spatio-textual recommender system," *Expert Syst Appl*, vol. 78, pp. 396–406, Jul. 2017, doi: 10.1016/J.ESWA.2017.01.060.
- [9] R. Yeganegi, S. Haratizadeh, and M. Ebrahimi, "STAR: A session-based time-aware recommender system," *Neurocomputing*, vol. 573, 2024, doi: 10.1016/j.neucom.2023.127104.
- [10] E. Khazaei and A. Alimohammadi, "Context-aware group-oriented location recommendation in location-

- based social networks,” *ISPRS Int J Geoinf*, vol. 8, no. 9, 2019, doi: 10.3390/ijgi8090406.
- [11] W. Shafqat and Y. C. Byun, “A context-aware location recommendation system for tourists using hierarchical LSTM model,” *Sustainability (Switzerland)*, vol. 12, no. 10, 2020, doi: 10.3390/su12104107.
- [12] C. Fast and F. Labs, “Session-Based Recommender Systems,” no. May, 2021.
- [13] J. Zhang, C. Ma, X. Mu, P. Zhao, C. Zhong, and A. Ruhan, “Recurrent convolutional neural network for session-based recommendation,” *Neurocomputing*, vol. 437, 2021, doi: 10.1016/j.neucom.2021.01.041.
- [14] H. Yin and B. Cui, *Spatio-Temporal Recommendation in Social Media*. 2016. [Online]. Available: <http://link.springer.com/10.1007/978-981-10-0748-4>
- [15] C. P. R. Proceedings, “Contextualized Point-of-Interest Recommendation Contextualized Point-of-Interest Recommendation,” 2023, doi: 10.24963/ijcai.2020/344.
- [16] S. Safavi, M. Jalali, and M. Houshmand, “Toward Point-of-Interest Recommendation Systems: A Critical Review on Deep-learning Approaches,” *Electronics (Switzerland)*, vol. 11, no. 13, 2022, doi: 10.3390/electronics11131998.
- [17] Z. Khan, M. I. Hussain, N. Iltaf, J. Kim, and M. Jeon, “Contextual recommender system for E-commerce applications,” *Appl Soft Comput*, vol. 109, p. 107552, Sep. 2021, doi: 10.1016/J.ASOC.2021.107552.
- [18] D. Jannach, M. Ludewig, and L. Lerche, “Session-based item recommendation in e-commerce: on short-term intents, reminders, trends and discounts,” *User Model User-adapt Interact*, vol. 27, no. 3–5, 2017, doi: 10.1007/s11257-017-9194-1.
- [19] X. Fan, Y. Hu, Z. Zheng, Y. Wang, P. Brezillon, and W. Chen, “CASR-TSE: Context-Aware Web Services Recommendation for Modeling Weighted Temporal-Spatial Effectiveness,” *IEEE Trans Serv Comput*, p. 1, 2017, doi: 10.1109/TSC.2017.2782793.
- [20] 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*, in RecSys ’10. New York, NY, USA: ACM, 2010, pp. 79–86. doi: 10.1145/1864708.1864727.
- [21] S. Y. Jeong and Y. K. Kim, “Deep Learning-Based Context-Aware Recommender System Considering Contextual Features,” *Applied Sciences (Switzerland)*, vol. 12, no. 1, 2022, doi: 10.3390/app12010045.
- [22] S. Y. Jeong and Y. K. Kim, “Deep Learning-Based Context-Aware Recommender System Considering Change in Preference,” *Electronics (Switzerland)*, vol. 12, no. 10, 2023, doi: 10.3390/electronics12102337.
- [23] L. Page, S. Brin, R. Motwani, and T. Winograd, “The PageRank Citation Ranking: Bringing Order to the Web,” *World Wide Web Internet And Web Information Systems*, vol. 54, no. 1999–66, 1998, doi: 10.1.1.31.1768.
- [24] S. Park, W. Lee, B. Choe, and S. G. Lee, “A Survey on Personalized PageRank Computation Algorithms,” *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2952653.
- [25] F. Jiang and Z. Wang, “Pagerank-based collaborative filtering recommendation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010. doi: 10.1007/978-3-642-16167-4\_76.
- [26] P. T. Nguyen, P. Tomeo, T. Di Noia, and E. Di Sciascio, “An evaluation of simrank and personalized pagerank to build a recommender system for the web of data,” in *WWW 2015 Companion - Proceedings of the 24th International Conference on World Wide Web*, 2015. doi: 10.1145/2740908.2742141.
- [27] S. Chaudhari, A. Azaria, and T. Mitchell, “An entity graph based Recommender System,” *AI Communications*, vol. 30, no. 2, 2017, doi: 10.3233/AIC-170728.
- [28] S. Guo, W. Zhang, and S. Zhang, “A pagerank-based collaborative filtering recommendation approach in digital libraries,” *Tehnicky Vjesnik*, vol. 24, no. 4, 2017, doi: 10.17559/TV-20160602011232.
- [29] C. Musto, P. Lops, M. de Gemmis, and G. Semeraro, “Context-aware graph-based recommendations exploiting Personalized PageRank,” *Knowl Based Syst*, vol. 216, Mar. 2021, doi: 10.1016/j.knosys.2021.106806.
- [30] Y. Zheng, R. Burke, and B. Mobasher, “Splitting Approaches for Context-aware Recommendation: An Empirical Study,” in *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, in SAC ’14. New York, NY, USA: ACM, 2014, pp. 274–279. doi: 10.1145/2554850.2554989.
- [31] 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)*, in ICDMW ’15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 1668–1671. doi: 10.1109/ICDMW.2015.222.
- [32] L. Baltrunas, B. Ludwig, and F. Ricci, “Matrix Factorization Techniques for Context Aware Recommendation,” in *Proceedings of the Fifth ACM Conference on Recommender Systems*, in RecSys ’11. New York, NY, USA: ACM, 2011, pp. 301–304. doi: 10.1145/2043932.2043988.

## A GRAPH-BASED SPREADING ACTIVATION METHOD FOR CONTEXT-AWARE COLLABORATIVE RECOMMENDER SYSTEMS

**Abstract:** Context-aware recommender systems are specifically designed to leverage contextual information in the recommendation process to better match items to users, achieving higher effectiveness compared to traditional systems that do not use context. A major challenge in building recommender systems is data sparsity, which affects recommendation quality. While incorporating context provides valuable information, it also increases data sparsity and computational complexity. In this study, I will propose a new approach for context-aware recommender systems to comprehensively integrate contextual information, minimize the impact of data sparsity, and maintain low computational complexity. This is achieved by combining user splitting based on context with a graph-based spreading activation method. Experiments on several datasets show that the proposed approach significantly improves prediction accuracy and is

feasible in terms of time for practical deployment compared to previous baseline context-aware collaborative methods.

**Keyword:** Context-Aware Recommender Systems – CARS; Spreading Activation; Context-Aware Collaborative Filtering – CACF; User Splitting; Context Awareness.



**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, hệ tư vấn, khai phá dữ liệu, phát triển ứng dụng đa phương tiện.