**ỦY BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**VÕ ĐĂNG QUANG**

**KHẢO SÁT CÁC HỆ THỐNG KHUYẾN NGHỊ TRÊN MỘT SỐ TRANG THƯƠNG MẠI ĐIỆN TỬ**

**ĐỀ CƯƠNG ĐỒ ÁN CHUYÊN NGÀNH**

NGÀNH: CÔNG NGHỆ THÔNG TIN

**Thành phố Hồ Chí Minh, tháng 12 năm 2023**

**ỦY BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**VÕ ĐĂNG QUANG**

**KHẢO SÁT CÁC HỆ THỐNG KHUYẾN NGHỊ TRÊN MỘT SỐ TRANG THƯƠNG MẠI ĐIỆN TỬ**

**ĐỀ CƯƠNG ĐỒ ÁN CHUYÊN NGÀNH**

ngành: CÔNG NGHỆ THÔNG TIN

**Giảng viên phụ trách**

**TS. PHAN TẤN QUỐC**

**Thành phố Hồ Chí Minh, tháng 12 năm 2023**

# **Lời cam đoan**

Em tên là Võ Đăng Quang, em xin cam đoan rằng đồ án chuyên ngành “*Khảo sát các hệ thống khuyến nghị trên một số trang thương mại điện tử*” là công trình nghiên cứu của riêng em dưới sự hướng dẫn trực tiếp của TS. Phan Tấn Quốc.

Mọi trích dẫn sử dụng trong báo cáo đều được ghi rõ nguồn tài liệu tham khảo theo đúng quy định.

Em xin hoàn toàn chịu trách nhiệm và chịu mọi hình thức kỷ luật theo quy định nếu có bất kì hành vi vi phạm, gian trá nào.

**Thành phố Hồ Chí Minh, tháng 12 năm 2023**

**Võ Đăng Quang**

# **Lời cảm ơn**

Trước hết em xin gửi đến lời cảm ơn chân thành và sâu sắc nhất đến thầy TS. Phan Tấn Quốc, người trực tiếp hướng dẫn và tận tình chỉ bảo cho em cho tới khi em hoàn thành đồ án của mình.

Tiếp đến em xin giành lời cảm ơn đến quý thầy cô khoa Công nghệ thông tin – Trường Đại học Sài Gòn đã truyền đạt cho em những kiến thức vô cùng quý báu và bổ ích trong suốt quá trình nghiên cứu và học tập tại trường.

Xin chân thành cảm ơn tới những người bạn đã luôn sát cánh cùng em, những lời động viên, những lần hỗ trợ những lúc cần thiết đã phần nào giúp em hoàn thành đồ án này.

Cuối cùng, em xin cảm ơn đến ba mẹ và người thân trong gia đình đã hỗ trợ và tạo điều kiện thuận lợi cho em trong suốt thời gian học tập và nghiên cứu tại Đại học Sài Gòn.

# **Mục lục**

[**Lời cam đoan** i](#_Toc148302733)

[**Lời cảm ơn** ii](#_Toc148302734)

[**Mục lục** iii](#_Toc148302735)

[**Danh mục hình ảnh** v](#_Toc148302736)

[**Lời mở đầu** 1](#_Toc148302737)

[**Chương 1. TỔNG QUAN VỀ HỆ THỐNG KHUYẾN NGHỊ** 3](#_Toc148302738)

[**1.1.** **Tổng quan về hệ thống khuyến nghị** 3](#_Toc148302739)

[**1.1.1.** **Khái niệm** 3](#_Toc148302740)

[**1.1.2.** **Thành phần của hệ thống khuyến nghị** 4](#_Toc148302741)

[**1.1.3.** **Mục đích xây dựng hệ thống cho các trang thương mại điện tử** 5](#_Toc148302742)

[**1.2.** **Bài toán khuyến nghị** 5](#_Toc148302743)

[**1.3.** **Cách tiếp cận của bài toán khuyến nghị** 7](#_Toc148302744)

[**1.3.1.** **Tiếp cận lọc dựa trên nội dung** 7](#_Toc148302745)

[**1.3.1.1.** **Ưu điểm của cách tiếp cận này** 9](#_Toc148302746)

[**1.3.1.2.** **Nhược điểm của cách tiếp cận này** 10](#_Toc148302747)

[**1.3.2.** **Tiếp cận lọc cộng tác** 10](#_Toc148302748)

[**1.3.2.1.** **Lọc cộng tác dựa trên bộ nhớ** 11](#_Toc148302749)

[**1.3.2.2.** **Lọc cộng tác dựa trên mô hình** 13](#_Toc148302750)

[**1.3.2.3.** **Ưu điểm và nhược điểm** 13](#_Toc148302751)

[**1.3.3.** **Tiếp cận lọc lai ghép** 14](#_Toc148302752)

[**1.3.3.1.** **Lai có trọng số** 14](#_Toc148302753)

[**1.3.3.2.** **Lai chuyển đổi** 15](#_Toc148302754)

[**1.3.3.3.** **Lai trộn** 15](#_Toc148302755)

[**1.3.3.4.** **Lai kết hợp đặc trưng** 16](#_Toc148302756)

[**1.3.3.5.** **Lai theo đợt** 16](#_Toc148302757)

[**1.3.3.6.** **Lai tăng cường đặc trưng** 18](#_Toc148302758)

[**1.3.3.7.** **Lai meta** 19](#_Toc148302759)

[**1.4.** **Kết chương** 19](#_Toc148302760)

[**Tài liệu tham khảo** 21](#_Toc148302761)

# **Danh mục hình ảnh**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Hình | Tên | Trang |
| 1 | 0.1 | Nhân viên tư vấn sản phẩm cho khách hàng | 1 |
| 2 | 1.1.1 | Shopee khuyến nghị sản phẩm dựa trên lịch sử tìm kiếm | 3 |
| 3 | 1.1.2 | Các thành phần của hệ thống khuyến nghị | 4 |
| 4 | 1.3.1-1 | Mô hình hệ thống khuyến nghị | 7 |
| 5 | 1.3.1-2 | Kiến trúc tổng quan của tiếp cận nội dung | 8 |
| 6 | 1.3.2 | Ma trận tiếp cận lọc cộng tác | 11 |
| 7 | 1.3.2.1 | Mô hình tiếp cận lọc cộng tác dựa trên bộ nhớ | 12 |
| 8 | 1.3.3 | Mô hình lọc lai ghép | 14 |
| 9 | 1.3.3.5 | Mô hình nhà hàng Entree Chicago | 17 |
| 10 | 1.3.3.6 | Mô hình hệ thống LIBRA | 18 |

# **Lời mở đầu**

Thương mại điện tử từ nhưng năm 2018 trở lại đây như một quả bom bùng nổ trên khắp thế giới, nó đã trở thành lĩnh vực có ảnh hưởng cực kì quan trọng đến kinh tế của nhiều quốc gia. Sự phát triển của thương mại điện tử đã giúp các doanh nghiệp đạt hiệu quả cao trong kinh doanh dẫn đến việc kinh doanh điện tử (E-business) ra đời.

Tư vấn, khuyến nghị sản phẩm và dịch vụ chăm sóc khách hàng rất quan trọng bởi nó tác động trực tiếp tới doanh thu và lợi nhuận của các doanh nghiệp. Nhờ có thương mại điện tử, vô số hình thức kinh doanh mới được hình thành, nổi bật nhất là hình thức mua bán sản phẩm trực tuyến. Với hình thức mới này, thương mại điện tử cho phép các công ty, doanh nghiệp cung cấp cho khách hàng nhiều sự lựa chọn trong cách mua hàng và số lượng sản phẩm. Tuy nhiên tỉ lệ thuận với sự đa dạng này là chính là lượng thông tin khổng lồ mà khách hàng phải xử lý trước khi muốn bỏ món hàng nào đó vào giỏ hàng. Và để giải quyết vấn đề này cũng như tăng cường trải nghiệm của khách hàng trên trang thương mại điện tử của mình, các doanh nghiệp đã áp dụng hệ thống khuyến nghị (Recommender System – RS) như một phần quan trọng không thể thiếu trong hệ thống của mình.

Hệ thống khuyến nghị đóng vai trò như một nhân viên hỗ trợ khách hàng đưa ra các quyết định mua hàng đúng đắn.



*Hình 0.1. Nhân viên tư vấn sản phẩm cho khách hàng*

Bằng cách xác định mục đích, nhu cầu và sở thích của khách hàng, hệ thống có thể đưa ra các gợi ý giúp người dùng có thể nhanh chóng chọn được sản phẩm theo ý của mình.

Những nghiên cứu về tính hiệu quả của hệ thống khuyến nghị sẽ được giải thích vì sao các doanh nghiệp áp dụng nó cho hệ thống của mình. Vì vậy, để tìm hiểu rõ hơn về cách hoạt động của hệ thống khuyến nghị cũng như cách các trang thương mại điện tử sử dụng hệ thống khuyến, em quyết định chọn nghiên cứu đề tài: “Khảo sát các hệ thống khuyến nghị trên một số trang thương mại điện tử”.

Hệ thống khuyến nghị là một nhánh con của hệ thống lọc thông tin, nhằm tìm cách dự đoán đánh giá của người dùng sẽ đưa ra cho một sản phẩm. Trong nhiều năm, các phương pháp khuyến nghị khác nhau đã được đề xuất và có ba loại được sử dụng khá phổ biến hiện nay: Lọc dựa trên nội dung (Content-based-filtering), Lọc cộng tác (Collaborative Filtering) và kết hợp cả hai phương pháp trên ta có được phương pháp thứ ba đó là Lai ghép (Hybrid method).

Trong đề tài này, ba trang thương mại điện tử được khảo sát là Shopee, Lazada và Tiki. Quá trình nghiên cứu em sẽ trình bày rõ thế nào là một hệ thống khuyến nghị cũng như cấu trúc của nó. Kết hợp với việc quan sát trực quan trên các trang thương mại được kể trên và cho ra các nhận định của mình. Và cuối cùng, em sử dụng phương pháp so sánh để đánh giá hệ thống và đưa quan điểm, góp ý của mình về từng hệ thống khuyến nghị của ba đối được khảo sát.

Đồ án chuyên ngành của em được chia thành các phần như sau:

Chương 1: Trình bày tổng quan về khái niệm thế nào là hệ thống khuyến nghị. Đồng thời, em cũng trình bày rõ cấu trúc và các thành phần tạo nên một hệ thống khuyến nghị.

Chương 2: Trong chương này, em sẽ trình bày chi tiết hơn về các phương pháp khuyến nghị đã nêu ở trên.

Chương 3: Kết quả sau khi khảo sát hệ thống khuyến nghị trên một ba trang thương mại điện tử sẽ được trình bày ở đây và em cũng đưa ra đánh giá, so sánh của mình trong phần này.

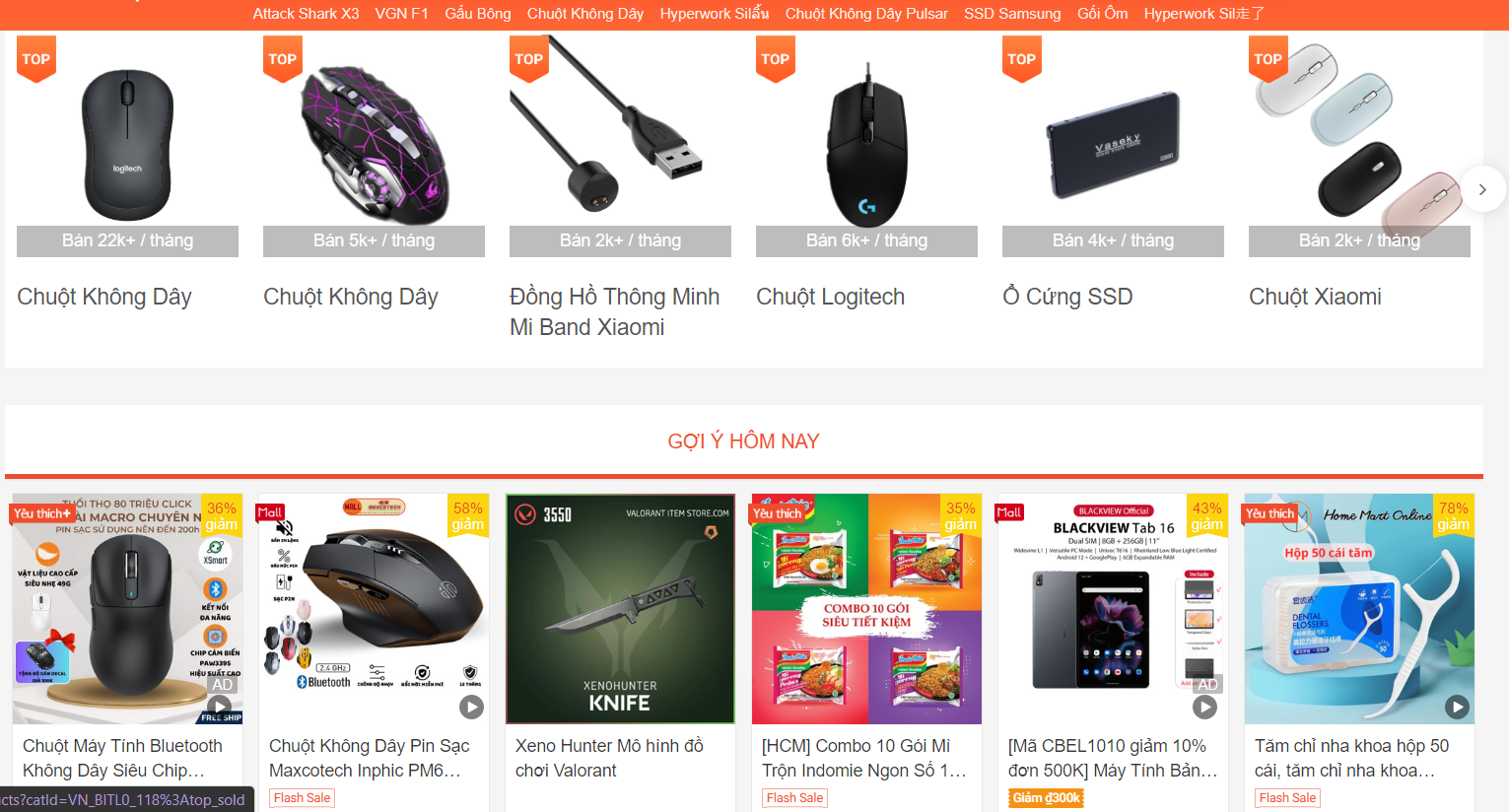
# **Chương 1. TỔNG QUAN VỀ HỆ THỐNG KHUYẾN NGHỊ**

## **Tổng quan về hệ thống khuyến nghị**

### **Khái niệm**

Hệ thống khuyến nghị là hệ thống lọc thông tin nhằm dự đoán sở thích, đánh giá, mối quan tâm, nhu cầu của khách hàng nhằm đưa ra một hoặc nhiều sản phẩm mà tỉ lệ cao khách hàng sẽ thích. Khác với các công cụ tìm kiếm (Search Engine), hệ thống khuyến nghị chủ động đưa ra các dự đoán mà không cần người dùng yêu cầu. Trong khi các công cụ tìm kiếm đòi hỏi người dùng phải mô tả yêu cầu của mình, ví dụ dưới dạng các từ khoá, tức là người dùng phải biết mình muốn gì, cũng như phải biết hệ thống có thể cung cấp thứ mình muốn thì hệ thống khuyến nghị không cần cả hai thông tin này [18] [19] [22] [25].

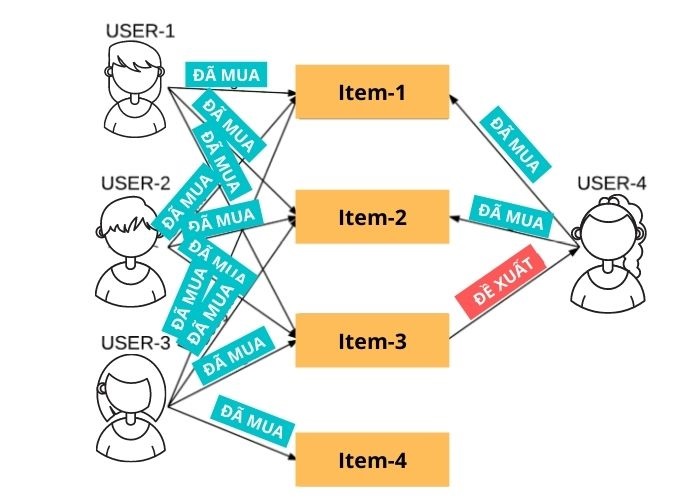
Để đơn giản hơn có thể hiểu rằng khi được tích hợp vào các hệ thống thương mại điện tử, hệ thống khuyến nghị sẽ giúp người dùng dễ dàng hơn trong quá trình tìm kiếm thông tin liên quan trở nên dễ dàng hơn vì chúng tự động tìm đến người dùng.



*Hình 1.1.1. Shopee khuyến nghị sản phẩm dựa trên lịch sử tìm kiếm*

Nguồn gốc của hệ thống khuyến nghị cũng xuất phát từ hành vi của người sắp mua một sản phẩm: họ thường hỏi bạn bè, người quen, chuyên gia hay những người tư vấn về sản phẩm đó [23]. Lúc này người được hỏi sẽ tiến hành thu thập thông tin về sở thích, nhu cầu, tài chính,… kết hợp với kiến thức của mình về sản phẩm đó và cho ra các đề xuất phù hợp nhất với các dữ liệu thu thập được.

### **Thành phần của hệ thống khuyến nghị**



*Hình 1.1.2. Các thành phần của hệ thống khuyến nghị*

Thứ nhất: Thành phần thứ nhất chúng ta phải quan tâm đó chính là người dùng (User). Rõ ràng đây là thành phần quan trọng không thể thiếu trong hệ thống khuyến nghị.

Thứ hai: Điều tiếp theo chúng ta cần phải quan tâm là mục tin (Items) vì các mục tin này là các sản phẩm trên các trang web bán hàng, một bài viết, một bài hát của các trang nghe nhạc. Có người dùng thì chắc chắn phải có các items để gợi ý.

Thứ ba: Thứ để hệ thống khuyến nghị hoạt động đó chính là phản hồi (Feedback) của người dùng, Feedback ở đây có thể là điểm đánh giá, bình luận thể hiện sự quan tâm của người dùng lên mỗi sản phẩm… Vì chúng ta phải định lượng các chỉ số phản hồi này thì mới có cơ sở để đưa ra gợi ý cho người dùng.

### **Mục đích xây dựng hệ thống cho các trang thương mại điện tử**

Mục đích của hệ thống khuyến nghị đó là:

* Gia tăng doanh số bán hàng: Đây có lẽ là lý do quan trọng nhất cho những trang thương mại điện tử, đơn giản là họ sẽ bán thêm được hàng so với những trang mà không sử dụng hệ thống khuyến nghị. Hệ thống đề xuất càng chính xác thì doanh số bán ra càng cao [22] [23].
* Thỏa mãn nhu cầu khách hàng: Thiết kế và cài đặt tốt sẽ cải thiện trải nghiệm người dùng. Người dùng sẽ thấy rằng trang thương mại điện tử này rất đáng tin, chính xác và cùng với một UI được thiết kế hợp lý, khách hàng sẽ muốn quay lại website [22] [25].
* Nắm được nhu cầu khách hàng: Một lí do quan trọng khác của hệ thống khuyến nghị. Nếu khách hàng không thích những thứ mà bạn đề xuất, bạn vẫn có thể thu thập lại được những thông tin này và tái sử dụng nó để cải thiện độ chính xác [23] [25].

Với mục đích như vậy, hệ thống khuyến nghị hiển nhiên trở thành một công cụ đắc lực cho các trang thương mại cạnh tranh với nhau.

## **Bài toán khuyến nghị**

Phát biểu bài toán [20] [24]:

**Input:**

* Cho tập người dùng U, với mỗi người dùng Ui ∈ U có các đặc điểm I = {I1, I2,…, Ik}.
* Cho tập các sản phẩm P, với mỗi sản phẩm Pj ∈ P có các điểm đặc trưng J = {J1, J2,..., Jx}.
* Một ma trận R = (Rij) với i = 1…N và j = 1... M, thể hiện mối quan hệ giữa tập người dùng U và tập sản phẩm P. Trong đó Rij là dánh giá của người dùng Ui cho sản phẩm Pj, N và M lần lượt là số lượng người dùng và số sản phẩm.

**Output:**

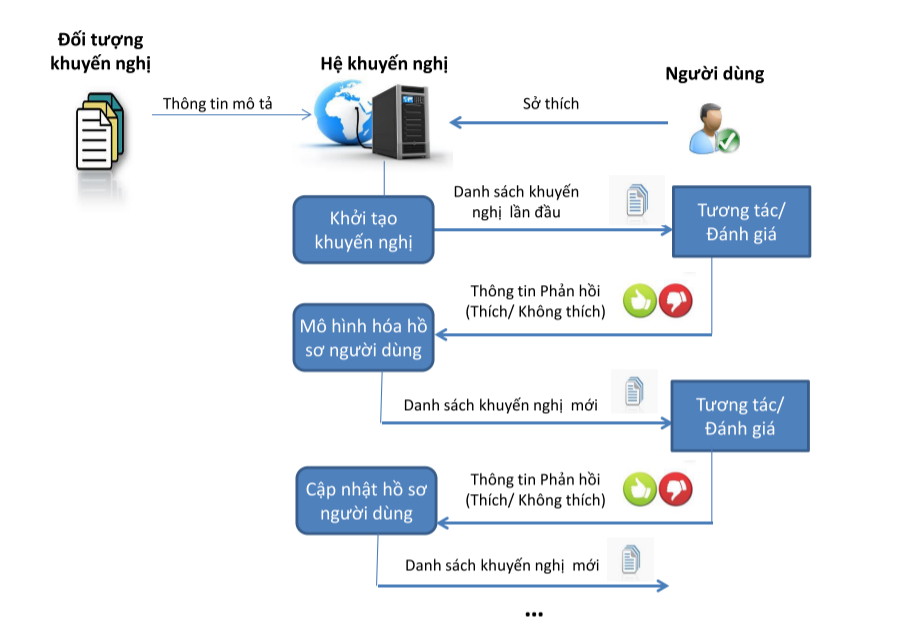
* Danh sách các sản phẩm Pj thuộc P có độ phù hợp với người dùng Ui nhất.
* Để giải bài toán này chúng ta cần xây dựng hàm *F(Ui, Pi)* để đo độ phù hợp sản phẩm Pi đối với người dùng Ui, từ đó sẽ lấy ra được danh sách các sản phẩm phù hợp với yêu cầu cao nhất.
* Tùy thuộc vào phương pháp sử dụng ta có nhiều cách xây dựng hàm F, các cách xây dụng hàm F phụ thuộc chủ yếu bởi các yếu tố sau:
  + Đặc điểm của người dùng Ui (lọc theo nội dung người dùng) : điều này được đánh giá chủ quan bởi các quy luật tự nhiên, hoặc các quy tắc cơ bàn. Ví dụ Ui là nam thì sẽ có xu hướng mua các sản phẩm dành cho nam hơn các sản phẩm dành cho nữ, Ui thuộc độ tuổi khoảng 15 tới 27 thường sẽ thích nghe những bản nhạc điện tử, nhạc rap...
  + Đặc điểm của sản phẩm Pj (lọc theo nội dung sản phẩm): giống như lọc theo nội dung người dùng, các sản phẩm có đặc điểm giống nhau, thì cũng có khá nãng được người dùng đánh giá như nhau. Ví dụ về đặc điểm của các bài hát có thể là thể loại âm nhạc, nhạc chủ đề của phim gì,...
  + Lịch sử giao dịch của người dùng Ui: từ lịch sử giao dịch có thể suy ra chủ đề mà Ui quan tâm, do đó những sản phẩm có cùng lĩnh vực sẽ có độ liên quan cao hơn. Ví dụ một người đã từng mua điện thoại và máy tính bảng thì có thể dự đoán người này yêu công nghệ. Từ đó suy ra người này sẽ có khá năng mua các sản phẩm công nghệ cao hơn các sản phẩm khác.
  + Những người dùng Ut khác có cùng các đặc điểm giống Ui: với quan niệm rằng những người dùng giống nhau sẽ thích, đánh giá những sản phẩm giống nhau. Các đặc điểm của Ut bao gồm tập đặc điểm I ban đầu, kết hợp với các đặc điểm cộng tác như cùng mua mặt hàng nào đó, có các hành vi giống nhau… Việc tìm hiểu những mặt hàng mà Ut đã từng quan tâm sẽ đưa ra được gợi ý phù hợp cho người dùng Ui.

Việc xây dựng hàm F và ước lượng giá trị P với người dùng U có thể được thực hiện bằng nhiều phương pháp khác nhau [20]. Phần tiếp theo sẽ trình bày sơ lược về những cách tiếp cận đó thông qua các nghiên cứu đi trước

## **Cách tiếp cận của bài toán khuyến nghị**

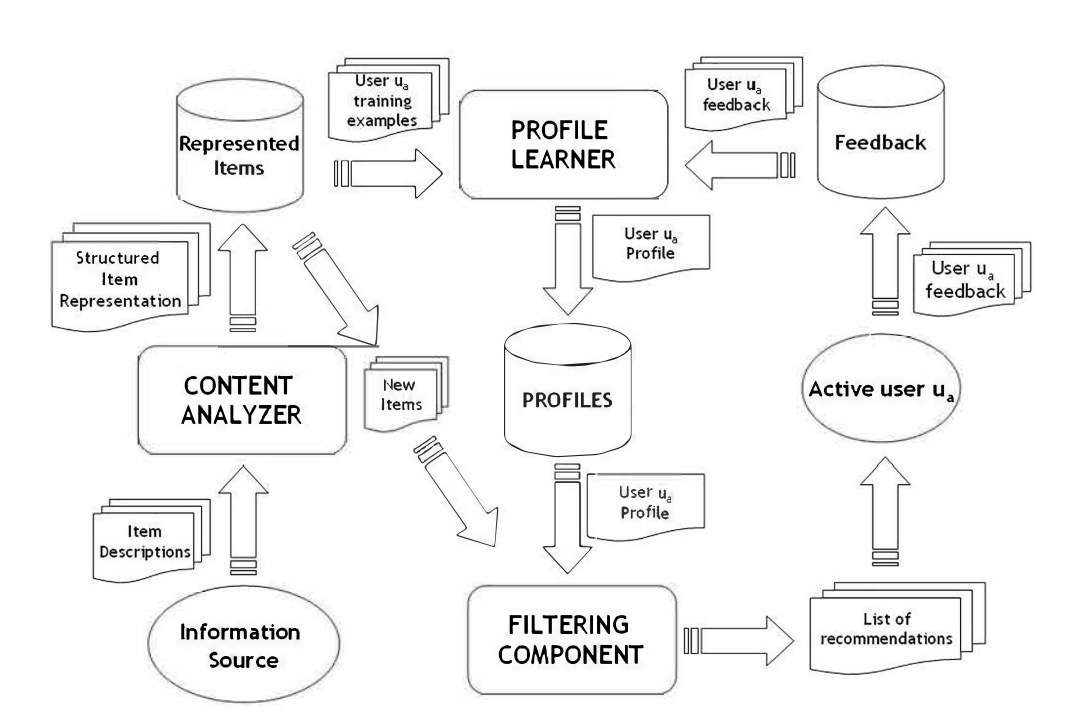
### **Tiếp cận lọc dựa trên nội dung**

Theo Bobadilla và cộng sự [6], người dùng Ui sẽ có hồ sơ người dùng U biểu diễn mức độ ưa thích với item là Pj ∈ P và giúp hệ khuyến nghị tiên đoán item đó có hữu ích không và mức độ hữu ích đối với Ui là như thế nào.



*Hình 1.3.1-1. Mô hình hệ thống khuyến nghị*

Tiếp cận dựa trên nội dung có nguồn gốc từ cộng đồng nghiên cứu về truy vấn thông tin và lọc thông tin [4] [5]. Cách tiếp cận này sẽ đưa ra các khuyến nghị với những đối tượng mà người dùng đã từng quan tâm trong quá khứ. Điển hình các trang thương mại sẽ đưa ra các sản phẩm liên quan đến chuột máy tính hay bàn phím nếu như trong quá khứ người dùng thường tìm kiếm các sản phẩm có chứa từ chuột và bàn phím.



*Hình 1. 3. 1-2. Kiến trúc tổng quan của tiếp cận nội dung*

Để dự đoán có hay không người dùng U sẽ thích đối tượng khuyến nghị P và thích nhiều hay ít, các phương pháp dựa trên tiếp cận nội dung thường sẽ thực hiện các bước sau [2]:

* Bước 1: Biểu diễn nội dung đối tượng khuyến nghị Pi ∈ P. Kí hiệu *UserProfile(U)*.
* Bước 2: Mô hình hoá sở thích người dùng Ui ∈ U. Kí hiệu *Item(P).*
* Bước 3: Ước lượng giá trị hữu ích dựa trên độ tương tự nội dung của đối tượng khuyến nghị P với hồ sơ người dùng U. Hệ thống sẽ ưu tiên khuyến nghị những đối tượng có nội dung tương tự cao so với hồ sơ người dùng U.

*F(Ui, Pi) = Sim(UserProfile(U), Item(P))*

Pasquale Lops và cộng sự đã tiến hành khảo sát, phân tích các hệ khuyến nghị dựa trên tiếp cận nội dung [14]. Theo Lops, hệ khuyến nghị dựa trên tiếp cận nội dung thường sẽ thực hiện ba việc chính và được đảm nhận bởi ba thành phần tương ứng:

Phân tích nội dung (Content Analyzer): Thành phần này có nhiệm vụ phân tích mô hình háo nội dung của các đối tượng khuyến nghị. Tuỳ vào bài toán, các phương pháp rút trích đặc trưng sẽ được dùng để chuyển đổi nội dung đối tượng khuyến nghị từ định dạng gốc sang không gian đặc trưng.

Mô hình hoá hồ sơ người dùng (Profile Learner): Các nghiên cứu thường dùng các phương pháp học máy giám sát để học hồ sơ người dùng dựa trên đặc trưng của các đối tượng mà người dùng thích hay không thích trong quá khứ. Qua thời gian sở thích người dùng có thể thay đổi, hệ thống thường sẽ định kì học và cập nhật lại hồ sơ người dùng.

Lọc nội dung (Filtering Component): Thành phần này có nhiệm vụ so khớp hồ sơ người dùng với nội dung của các đối tượng để thực hiện khuyến nghị những đối tượng phù hợp với sở thích người dùng.

#### **Ưu điểm của cách tiếp cận này**

Michael Pazzani và đồng nghiệp đã dùng tiếp cận dựa trên mô hình để phát triển hệ thống khuyến nghị các trang web phù hợp với sở thích người dùng [96]; Trong bài báo cáo [13], Jiahui Liu và cộng sự đã giới thiệu việc phát triển hệ thống khuyến nghị tin tức Google News. Hệ thống đã phân tích tập tin log để tìm hiểu sở thích, hành vi đọc tin của người dùng. Kết quả phân tích, Jiahui Liu đã dùng cách tiếp cận dựa trên nội dung để thực hiện khuyến nghị tin tức cho người dùng. Cả hai nghiên cứu trên đều cho ra ưu điểm như sau:

Một là, hệ thống không yêu cầu cần có quá nhiều dữ liệu từ những người dùng khác để đạt được độ chính xác khuyến nghị chấp nhận được. Đối với phương pháp lọc dựa trên cộng tác hệ thống RS cần phải phân tích toàn bộ dữ liệu tương tác để tìm ra các quy luật (là những sản phẩm được người dùng ưa thích) thì mới có thể đưa ra gợi ý, tuy nhiên với phương pháp lọc dựa trên nội dung thì hệ thống chỉ dựa trên thông tin nội dung sản phẩm và dữ liệu tương tác của từng người dùng cụ thể để đưa ra gợi ý.

Hai là, có thể đưa ra khuyến nghị về các sản phẩm mới dựa trên thông tin mô tả có sẵn mà không cần phải có dữ liệu tương tác từ người dùng. Điều này xuất phát từ cơ chế hoạt động của lọc nội dung dựa trên phân tích về nội dung cấu tạo nên sản phẩm để đưa ra gợi ý, do đó khi một sản phẩm mới xuất hiện trong hệ thống thì chỉ cần nội dung của sản phẩm này “tương tự” với những sản phẩm trước đó đã được người dùng ưa thích sẽ được hệ thống khuyến nghị cho người dùng.

Ba là, trong trường hợp người dùng có những sở thích đặc thù dựa trên nội dung của sản phẩm nhưng những sản phẩm này lại không phổ biến trong dữ liệu ma trận tương tác thì hệ thống vẫn có khả năng khám phá được nhờ thông tin nội dung.

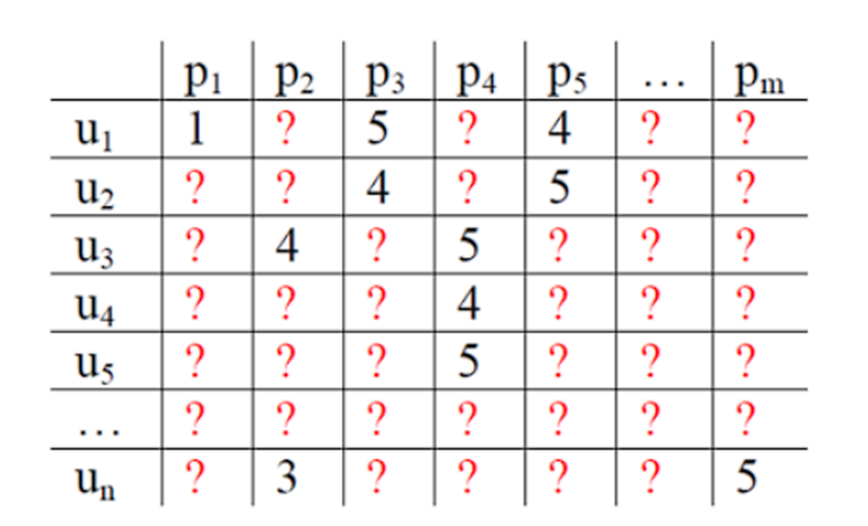
#### **Nhược điểm của cách tiếp cận này**

Bên cạnh ưu điểm thì tiếp cận lọc dựa trên nội dung có một số hạn chế có thể kể đến như sau:

* Hạn chế về phân tích nội dung: Hệ thống sẽ không thể phân biệt được đâu là tốt hay xấu, uy tín hay không để khuyến nghị. Bên cạnh đó các dữ liệu khác không phải là văn bản như hình ảnh, video, âm thanh,… cũng sẽ là rào cản lớn cho hướng tiếp cận này.
* Bên ngoài lĩnh vực quan sát: Người dùng chỉ được khuyến nghị các đối tượng mà tương tự cao với những gì mà hệ thống học được từ hồ sơ người dùng. Khi vượt quá phạm vi giám sát thì hệ thống không thể thực hiện khuyến nghị.
* Người dùng mới: Hệ thống khuyến nghị nội dung sẽ không thực hiện được cho những người dùng chưa có thông tin đánh giá trước đó. Nói cách khác, hệ thống không biết thật sự sở thích của người dùng đó là gì.

### **Tiếp cận lọc cộng tác**

Với các hạn chế mà hướng tiếp cận nội dung đang có thì tiếp cận lọc cộng tác sẽ giải quyết được. Tiếp cận lọc cộng tác [12] [17] được xem là cách tiếp cận thành công nhất để xây dựng hệ thống khuyến nghị. Ý tưởng của cách tiếp cận này là khai thác thông tin, hành vi quá khứ của người dùng dựa trên các đánh giá sẵn có từ ma trận đánh giá để tiên đoán, lượng hoá mức độ hữu ích của các đối tượng khuyến nghị mà người dùng chưa biết.

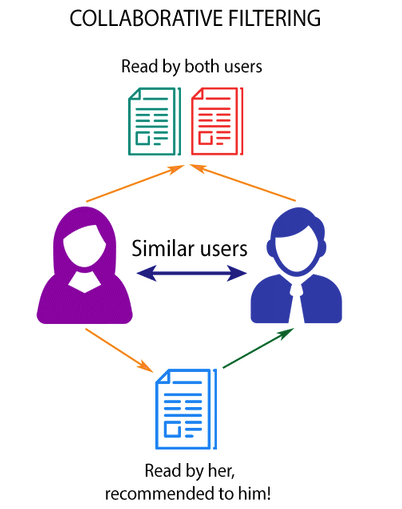


*Hình 1. 3. 2. Ma trận tiếp cận lọc cộng tác*

Trong phần tiếp sẽ trình bày về hai nhóm phương pháp chính của lọc cộng tác và một số nghiên cứu có liên quan.

#### **Lọc cộng tác dựa trên bộ nhớ**

Các hệ thống lọc cộng tác dựa trên bộ nhớ thường dùng các kĩ thuật thống kê để tìm kiếm những người dùng hoặc các đối tượng khuyến nghị tương tự nhau sau đó sẽ tìm cách ước lượng giá trị xếp hạng của người dùng trong ma trận. Phương pháp lọc cộng tác với cách tiếp cận dựa trên bộ nhớ có đặc trưng cơ bản là thường sử dụng toàn bộ dữ liệu đã có để dự đoán đánh giá của một người dùng nào đó về sản phẩm mới. Cách tiếp cận dựa trên bộ nhớ thường được chia thành: dựa trên người dùng và dựa trên sản phẩm. Nhưng về cơ bản, thì các thuật toán, kỹ thuật tỉnh toán cho lọc cộng tác dựa trên người dùng và lọc dựa trên đối tượng khuyến nghị từ ma trận đánh giá là tương tự nhau. Có khác chăng là kích thước của không gian người dùng và không gian đối tượng khuyến nghị sẽ ảnh hưởng đến tốc độ tính toán khi xác định nhóm các đối tượng tương tự.



*Hình 1. 3. 2. 1. Mô hình tiếp cận lọc cộng tác dựa trên bộ nhớ*

**Dựa trên người dùng**

Phương pháp này gồm 2 bước như sau:

* Bước 1: Tìm kiếm những người dùng có đánh giá tương tự với người dùng cần được dự đoán.
* Bước 2: Sử dụng đánh giá từ những người dùng được tìm thấy ở bước 1 để tính toán dự đoán cho người cần được dự đoán.

**Dựa trên sản phẩm**

Phương pháp này cũng gồm 2 bước:

* Bước 1: Xây dựng một ma trận để xác định mối quan hệ giữa các cặp sản phẩm với nhau.
* Bước 2: Kiểm tra thị hiếu của người dùng cần dự đoán bằng cách kiểm tra ma trận và kết hợp dữ liệu của người dùng đó.

#### **Lọc cộng tác dựa trên mô hình**

Phương pháp lọc cộng tác với cách tiếp cận dựa trên mô hình chủ yếu phát triển các mô hình bằng cách sử dụng các khai phá dữ liệu khác nhau, các thuật toán học máy để dự đoán đánh giá của người dùng về các mặt hàng chưa được đánh giá. Theo quan điểm xác xuất, thì các thuật toán lọc cộng tác dựa trên mô hình cần tính toán xác suất mà người dùng U đánh giá au,p cho một đối tượng khuyến nghị p, P(au,p|u, p). Quá trình đó có thể xem như việc tính toán giá trị kỳ vọng cho đánh giá của người dùng U với đối tượng khuyến nghị P [8].

Khác với lọc cộng tác dựa trên bộ nhớ, các thuật toán lọc cộng tác dựa trên mô hình dùng tập các đánh giá có sẵn trong ma trận A để học một mô hình đánh giá cho mỗi người dùng. Sau đó, mô hình học được sẽ dùng để tiên đoán các đánh giá khác. Một số thuật toán lọc cộng tác dựa trên mô hình được sử dụng phổ biến như Thuật toán lọc cộng tác gom cụm, Thuật toán lọc cộng tác dựa trên xác xuất Bayes [8], Thừa số hóa ma trận (Matrix Factorization)…

#### **Ưu điểm và nhược điểm**

**Ưu điểm:** Theo các tác giả Su và Khoshgoftaar, tiếp cận lọc cộng tác được xem là một trong những cách tiếp cận thành công nhất để thiết kế các thuật toán và xây dựng các hệ khuyến nghị [17]. Các ưu điểm của tiếp cận này là:

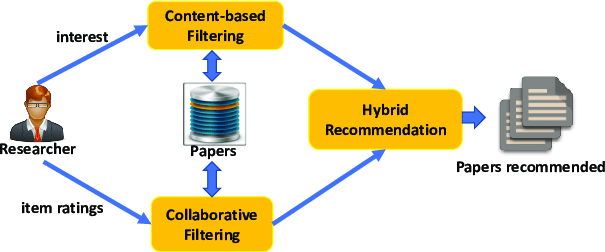
* Có khả năng dự doán sở thích và nhu cầu của người dùng để đưa ra các gợi ý sản phẩm phù hợp với từng khách hàng mà không cần hiểu sản phẩm.
* Phù hợp với những hệ thống lớn có nhiều đánh giá từ phía người dùng.

**Nhược điểm:** Tuy nhiên, theo nghiên cứu tổng quan của các tác giả Adomavicius và Tuzhilin [2]; Tác giả Bobadilla và cộng sự [6] thì tiếp cận lọc cộng tác cũng có một số hạn chế như sau:

* Ma trận thưa: Thực tế, không gian người dùng và đối tượng khuyến nghị là rất lớn. Trong khi đó, đánh giá của người dùng với các đối tượng khuyến nghị thực chất không nhiều. Điều đó ảnh hưởng đến độ chính xác khi khuyến nghị.
* Người dùng mới: Nhược điểm này cũng giống như hướng tiếp cận dựa trên nội dung. Hệ thống không thể biết được sở thích của người dùng nếu người dùng đó chưa hoặc có rất ít thông tin và đánh giá của họ.

### **Tiếp cận lọc lai ghép**

Nhằm kết hợp ưu điểm của cả hai phương thức khuyến nghị dựa trên lọc cộng tác và khuyến nghị dựa trên lọc nội dung cũng như hạn chế điểm yếu, Các nghiên cứu đi trước đã tập trung phát triển và đã tìm ra cách kết hợp kết quả gợi ý từ cả hai phương pháp tiếp cận nêu trên, nó được gọi là tiếp cận lại (Hybrid Approach).



*Hình 1. 3. 3. Mô hình lọc lai ghép*

Robin Burke đã khảo sát các phương pháp lai cho hệ khuyến nghị và trình bày 7 phương pháp khác nhau [9] [20].

#### **Lai có trọng số**

Hệ thống sử dụng giá trị xếp hạng từ thuật toán lọc theo nội dung và giá trị xếp hạng từ thuật toán lọc cộng tác để kết hợp lại thành một giá trị xếp hạng duy nhất bằng cách tính tổng theo trọng số đối với từng phương pháp. Claypool và cộng sự [10] đã xây dựng hệ thống lọc tin tức trực tuyến sử dụng khuyến nghị dựa trên nội dung và khuyến nghị dựa trên cộng tác. Kết quả cho ra trọng số ngang nhau, sau đó hệ thống sẽ dần hiệu chỉnh trọng số khi nhận được những đánh giá phản hồi từ người dùng. Tuy nhiên trọng số sẽ được điều chỉnh theo kết quả khuyến nghị của bộ phận nào đưa ra là chính xác hoặc không chính xác.

Ưu điểm và nhược điểm của lai có trọng số:

Ưu điểm: Tất cả khả năng, phương pháp khác nhau của hệ thống được tham gia vào quá trình khuyến nghị một cách minh bạch, tự nhiên và dễ dàng thực hiện, dễ dàng điều chỉnh.

Nhược điểm: Việc ước lượng trọng số lớn hay nhỏ cho phù hợp với những phương pháp khác nhau.

#### **Lai chuyển đổi**

Một số nghiên cứu đã thực hiện nghiên cứu một hệ thống khuyến nghị sử dụng phương pháp lai chuyển đổi giữa tiếp cận nội dung và lọc cộng tác. Hệ thống khuyến nghị thuộc nhóm lai chuyển đổi thường sử dụng một số điều kiện để chuyển đổi qua lại giữa các phương pháp khuyến nghị khác nhau. Ví dụ như hệ thống DailyLearner [3] của Billsus và Pazzani. Đây là một hệ thống gợi ý tin tức cho người dùng, các tác giả đã áp dụng phương pháp lọc nội dung trước, sau đó nếu như tiếp cận nội dung không thể thực hiện được thì tiếp cận lọc cộng tác sẽ được áp dụng.

Ưu điểm và nhược điểm của lai chuyển đổi:

Ưu điểm: Phương pháp tiếp cận này rất “nhạy” với các điểm mạnh và điểm yếu của các phương pháp khác nhau.

Nhược điểm: Tuy “nhạy” với điểm mạnh và điểm yếu của các phương pháp khác nhau, nhưng lai chuyển đổi yêu cầu cần phải xác định điều kiện để chuyển đổi giữa các phương pháp. Điều đó làm quá trình chuyển đổi trở nên phức tạp hơn.

#### **Lai trộn**

Tiếp cận lai trộn là phương pháp thực hiện các phương pháp khuyến nghị khác nhau một cách độc lập và kết hợp các kết quả từ phương pháp sẽ được chuyển thành danh sách đề xuất và được chuyển đến cho người dùng. Tiếp cận lai trộn có thể tránh được vấn đề người dùng mới (Khởi động lạnh – Cold Start). Các tác giả Smyth và Cotter dùng cách tiếp cận này để phát triển một hệ thống khuyến nghị chương trình truyền hình mỗi ngày thông qua Internet [16].

Cách tiếp cận này nhìn chung được đánh giá khá cao tuy nhiên vẫn còn vài nhược điểm:

Ưu điểm: Có thể giúp đề xuất các đối tượng tiềm năng mà bản thân một phương pháp riêng biệt không thể xác định được. Trộn lọc nội dung và lọc cộng tác có thể giúp giải quyết được vấn đề khởi động lạnh (Cold Start) và cũng có thể đa dạng hóa khuyến nghị.

Nhược điểm: Vì tiếp cận này sử dụng nhiều đề xuất từ các phương pháp khác nhau. Do đó hệ thống cần được xử lý, lọc các đề xuất đụng độ, trùng lắp từ các phương pháp khác nhau.

#### **Lai kết hợp đặc trưng**

Lai kết hợp đặc trưng là tiếp cận phát triển phương pháp khuyến nghị bằng cách sử dụng kết hợp thông tin đánh giá của người dùng với nội dung của đối tượng khuyến nghị [24].

Ưu điểm và nhược điểm của lai kết hợp đặc trưng:

Ưu điểm: Lai kết hợp đặc trưng cho phép hệ thống xem xét dữ liệu cộng tác, nhưng không chỉ phụ thuộc duy nhất vào dữ liệu cộng tác trong ma trận đánh giá. Ngược lại, hệ thống cũng có được thông tin về sự tương tự vốn có giữa các đối tượng khuyến nghị (dựa trên đặc trưng nội dung) mà không bị ảnh hưởng bởi dữ liệu cộng tác.

Nhược điểm: Khó khăn trong việc xác định các đặc trưng cộng tác và đặc trưng nội dung phù hợp.

#### **Lai theo đợt**

Lai theo đợt là tiếp cận mà các phương pháp khuyến nghị khác nhau được lần lượt áp dụng theo một thứ tự ưu tiên được xác định trước tùy vào mỗi ứng dụng cụ thể. Ví dụ, phương pháp khuyến nghị thứ nhất sinh ra một danh sách xếp hạng các Ứng viên (danh sách thô). Tiếp đó, những phương pháp khác với độ ưu tiên thấp hơn sẽ được áp dụng để lọc lại danh sách thô này. Lai theo đợt giúp phương pháp thứ hai tránh những đối tượng có thể không bao giờ cần khuyến nghị vì những đối tượng này đã được lọc qua phương pháp thứ nhất. Đồng thời, các đối tượng được ưu tiên chọn với phương pháp thứ nhất sẽ được tinh lọc, chứ không bị loại bỏ thông qua phương pháp thứ hai [24].



*Hình 1. 3. 3. 5. Hệ thống nhà hàng Entree Chicago*

Ví dụ hệ thống khuyến nghị nhà hàng Entree Chicago [9] dựa trên hệ khuyến nghị tri thức FindMe để hiệu chỉnh kết quả đưa ra bởi thuật toán lọc cộng tác.

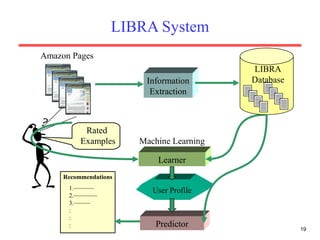
Ưu điểm và nhược điểm của lai theo đợt:

Ưu điểm: So với tiếp cận lại có trọng số (Weighted Hybrid) và một số tiếp cận lai khác thì việc lọc lại danh sách thô làm cho tiếp cận này hiệu quả hơn bởi vì các phương pháp tiếp theo chỉ thực hiện lọc trên một không gian nhỏ hơn, thay vì trên cả không gian tất cả các đối tượng khuyến nghị.

Nhược điểm: Khó khăn trong việc xác định độ ưu tiên giữa các phương pháp khác nhau cho mỗi ứng dụng cụ thể.

#### **Lai tăng cường đặc trưng**

Với tiếp cận lai tăng cường đặc trưng, phương pháp đầu sẽ học một mô hình để sinh ra đặc trưng tăng cường cho đầu vào của phương pháp tiếp theo. Tác giả Mooney và Roy đã giới thiệu một hệ thống thử nghiệm LIBRA (Learning Intelligent Book Recommending Agent) dùng cơ sở dữ liệu thông tin về sách được rút trích từ trang Amazon.com cho bài toán khuyến nghị sách [15].



*Hình 1. 3. 3. 6. Mô hình hệ thống libra*

Hệ thống này khai thác thông tin về những tác giả có liên quan, tiêu đề liên quan mà Amazon đã tạo ra dựa trên phương pháp lọc cộng tác. Sau đó, những thông tin này được dùng như những đặc trưng bổ sung thêm vào những đặc trưng nội dung để học hồ sơ sở thích người dùng.

Ưu điểm và nhược điểm của lai tăng cường đặc trưng:

Ưu điểm: Việc tăng cường đặc trưng dùng các phương pháp khác giúp hệ thống có thể cải tiến độ chính xác khuyến nghị mà không thay đổi, ảnh hưởng đến phương pháp khuyến nghị chính.

Nhược điểm: Khó khăn trong việc xác định đặc trưng tăng cường phù hợp.

#### **Lai meta**

Lai meta dùng mô hình được tạo ra bởi phương pháp trước làm đầu vào cho phương pháp sau. Với lại tăng cường đặc trưng (Feature Augmentation Hybird) thì phương pháp đầu sẽ học một mô hình để sinh ra đặc trưng làm đầu vào cho phương pháp tiếp theo. Trong khi lai meta thì cả mô hình của phương pháp thứ nhất sẽ làm đầu vào cho phương pháp thứ hai. Lai meta giữa lọc nội dung và lọc cộng tác phần nào giải quyết được vẫn đề ma trận thưa trong tiếp cận lọc cộng tác bởi vì lai meta sẽ tìm kiếm những người dùng tương tự dựa trên các đặc trưng nội dung trước khi áp dụng phương pháp lọc cộng tác. Đối với những người dùng có quá ít đánh giá thì việc xác định nhóm những người đồng sở thích thông qua lọc cộng khác sẽ không được chính xác.

Ưu điểm và nhược điểm của lai meta:

Ưu điểm: Với lai meta giữa lọc nội dung và cộng tác, phương pháp lọc cộng tác sẽ dễ dàng thực hiện tính toán trên “Dữ liệu dày” hơn so với dữ liệu thô trong ma trận đánh giá.

Nhược điểm: Khó khăn trong việc chọn phương pháp để thực hiện trước. Mỗi phương pháp được chọn vẫn phải gặp những hạn chế vốn có của nó.

Tóm lại, mỗi phương pháp tiếp cận lai đều có những ưu và nhược điểm vốn có của nó. Tiếp cận lai sẽ giúp giảm bớt phần nào hạn chế của các phương pháp khác nhau và đồng thời cũng được sử dụng cho nhiều trường hợp/mục đích khác nhau.

## **Kết chương**

Chương này đã trình bày tổng quan về thế nào là một hệ thống khuyến nghị, các thành phần của nó cũng như phát biểu bài toán khuyến nghị. Bên cạnh đó trong chương này cũng đã giới thiệu sơ lược các hướng tiếp cận của hệ thống khuyến nghị, đồng thời nêu ra các ưu điểm và nhược điểm của từng hướng tiếp cận.

Trên cơ sở kháo sát các nghiên cứu có liên quan, mục tiêu và phạm vi của đồ án là tập trung nghiên cứu các hướng tiếp cận hệ thống khuyến nghị đã được áp dụng trên các trang thương mại.

Với dữ liệu, thông tin thu thập được, đồ án này chọn tiếp tục khai thác sâu hơn về các vấn đề mà các trang thương mại đã được chọn khảo sát gặp phải. Chương tiếp theo sẽ trình bày rõ hơn các phương pháp tiếp cận phổ biến hiện nay, để rồi từ đó sẽ đưa ra các so sánh giữa các trang thương mại với nhau.

# **Tài liệu tham khảo**

[1] Adomavicius, G., Sankaranarayanan, R., Sen, S., and Tuzhilin, A., *Incorporating contextual information in recommender systems using a multidimensional approach*, New York University, NYC, 2005.

[2] Adomavicius, G., and Tuzhilin, A., *Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions*, New York University, NYC, 2005.

[3] Adomavicius, G., and Tuzhilin, A., *Context-aware recommender systems. In Recommender Systems Handbook*, New York University, NYC, 2011.

[4] Balabanovic, M. and Shoham, Y., *Fab: Content-based, collaborative recommendation*. ACM, 1997.

[5] Belkin, N. J. and Coft, W. B., *Information filtering and information retrieval: Two sides of the same coin?*. ACM, 1992.

[6] Bobadilla, J., Ortega, F.,Hernando, A., and Gutiérrez, A., *Recommender systems survey*. *Knowledge-Based Systems, 2013.*

[7] Baeza-Yates, R. A. and Ribeiro-Neto, B., *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA,1990.

[8] Breese, J. S., Heckerman, D., and Kadie, C., *Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, USA, 1998.

[9] Burke, R. D., *Hybrid recommender systems: Survey and experiments. User Modeling and User-adapted Interaction*, 2002.

[10] Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., and Sartin, M., *Combining content-based and collaborative filters in an online newspaper*, 1999.

[11] Francesco Ricci, Lior Rokach, Bracha Shapira, Paul B. Kantor, *Recommender Systems Handbook*, NewYork, DC: Springer, 2010.

[12] Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G., *Recommender Systems: An Introduction*, Cambridge University Press, New York, NY, USA, 1st edition, 2010.

[13] Liu, J., Dolan, P., and Pedersen, E. R., *Personalized news recommendation based on click behavior*, Cavazza, 2010.

[14] Lops, P., de Gemmis, M., and Semeraro, G., *Content-based recommender systems: State of the art and trends*. Springer, US, 2011.

[15] Mooney, R. J., and Roy, L., *Content-based book recommending using learning for text categorization*, NY, USA, 2000.

[16] Smyth, B. and Cotter, P., *A personalized television listing service*. ACM, 2000.

[17] Su, X. and Khoshgoftaar, T. M., *A survey of collaborative filtering techniques*, Adv. in Artif. Intell, 2009.

[18] Hồ Cẩm Hà, *Các hệ thống thông minh*, Hà Nội: Nhà xuất bản Đại học Sư Phạm, 2012.

[19] Kiều Xuân Chấn, “Nghiên cứu và xây dựng hệ thống khuyến nghị cho bài toán dịch giá trị gia tăng trong ngành viễn thông.”, Luận văn Thạc sĩ, Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội, Hà Nội, 2017.

[20] Trần Anh Dũng, “Ứng dụng các mô hình học sâu vào kĩ thuật lọc cộng tác dựa trên mô hình các hệ thống khuyến nghị thương mại.”, Luận văn Thạc sĩ, Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội, Hà Nội, 2020.

[21] Nguyễn Văn Đạt, “Giải quyết vấn đề phân phối trong hệ thống khuyến nghị dựa trên đặc trưng nội dung của đối tượng”, Luận văn Thạc sĩ, Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội, Hà Nội, 2021.

[22] Đỗ Thị Thanh Hà, “Phương pháp lọc cộng tác và ứng dụng trong hệ thống gợi ý bán hàng trực tuyến”, Luận văn Thạc sĩ, Trường Đại học Công nghệ thông tin và Truyền thông, Đại học Thái Nguyên, Thái Nguyên, 2020.

[23] Vũ Thị Nhạn, “Tổng hợp quan điểm trực tuyến của người tiêu dùng theo tính năng của sản phẩm”, Luận văn Thạc sĩ, Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội, Hà Nội, 2016.

[24] Huỳnh Ngọc Tín, “Phát triển một số phương pháp khuyến nghị hỗ trợ tìm kiếm thông tin học thuật dựa trên tiếp cận phân tích mạng xã hội”, Luận án tiến sĩ khoa học máy tính, Trường Đại học Công nghệ Thông tin, Đại học Quốc gia Thành phố Hồ Chí Minh, Thành phố Hồ Chí Minh, 2016.