

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



Vũ Chí Dũng

**Phát triển kỹ thuật tư vấn công bằng và ứng dụng trong
GIFT4Rec**

KHÓA LUẬN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY
Ngành: Công nghệ thông tin định hướng thị trường Nhật Bản

Hà Nội - 2023

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



Vũ Chí Dũng

**Phát triển kỹ thuật tư vấn công bằng và ứng dụng trong
GIFT4Rec**

KHÓA LUẬN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY
Ngành: Công nghệ thông tin định hướng thị trường Nhật Bản

Cán bộ hướng dẫn: PGS.TS. Hà Quang Thụy
ThS.Nguyễn Trần Ngọc Linh

Hà Nội - 2023

TÓM TẮT

Tóm tắt: Trong những năm gần đây, vấn đề người dùng khởi động nguội trong hệ tư vấn đang nhận được rất nhiều sự chú ý với nguyên nhân đến từ sự gia tăng nhanh chóng về số người dùng trong nền tảng công nghệ số của các công ty công nghệ lớn trên toàn cầu đang ứng dụng hệ tư vấn vào. Điều này thúc đẩy rất nhiều nghiên cứu ra đời, một trong số các lớp giải pháp chính là kết hợp thêm thông tin hỗ trợ của người dùng. Tuy nhiên, hầu hết các giải pháp hiện nay thường bỏ qua hiệu năng tư vấn trên người dùng đã có tương tác khi kết hợp thêm thông tin hỗ trợ. Thông tin hỗ trợ đôi khi không có ích trong việc đánh giá hành vi người dùng lại còn có thể gây nhiễu cho quá trình mô hình học và nắm bắt sở thích của nếu được kết hợp một cách không kiểm soát. Hậu quả xảy ra là hiệu năng tư vấn trên nhóm người dùng đã có tương tác giảm và gây ra tình trạng thiên vị hay thiếu công bằng trong hệ tư vấn. Cho đến hiện nay, theo như hiểu biết của tác giả khóa luận, mới có hai giải pháp xử lý vấn đề đặt ra. Trước tiên là mô hình mang tên DropoutNet [10] được đề xuất bởi Maksims Volkovs và cộng sự vào năm 2017 tại hội nghị NIPS. Cuối cùng là mô hình mang tên GIFT4Rec được đề xuất bởi chính tác giả khóa luận và các cộng sự khác tại hội nghị ACIIDS vào năm 2023. Khóa luận phân tích và nêu rõ các nhược điểm còn tồn tại trong GIFT4Rec khiến hiệu năng tư vấn còn bị giới hạn. Từ đây, khóa luận đề xuất phương thức khắc phục bằng việc cải tiến và áp dụng cách thức huấn luyện trong mô hình DropoutNet[10]. Bên cạnh đó, khóa luận cũng đề xuất thêm 1 giải pháp mới bắt nguồn từ lớp bài toán học liên tục để tăng hiệu quả học đa tác vụ trong GIFT4Rec cũng như tăng hiệu năng lẫn tính công bằng của hệ tư vấn. Kết quả thực nghiệm cho thấy sự vượt trội của mô hình đề xuất so với các giải pháp tốt nhất hiện có cũng như sự quan trọng của từng thành phần cải tiến.

Từ khóa: hệ tư vấn, khởi động nguội, công bằng

LỜI CẢM ƠN

Đầu tiên tôi xin cảm ơn PGS. TS Hà Quang Thùy và ThS Nguyễn Trần Ngọc Linh đã tận tình hướng dẫn và giúp đỡ tôi hoàn thiện khóa luận.

Tôi xin gửi lời cảm ơn chân thành tới các giảng viên hướng dẫn tại Trường Đại học Công nghệ - Đại học Quốc gia Hà Nội nói chung và Khoa Công nghệ Thông tin nói riêng đã chỉ dạy cho tôi những kiến thức bổ ích và vô giá trong suốt quãng thời gian trên giảng đường.

Tôi xin gửi lời cảm ơn chân thành tới chị Lê Hoàng Ngân cùng anh Hoàng Anh Dũng đã trực tiếp hỗ trợ và chỉ bảo cho tôi tận tình trong quá trình làm việc tại công ty. Tôi muốn gửi lời cảm ơn chân thành đến gia đình tôi, những người đã luôn trực tiếp ở bên động viên và chăm sóc tôi trong suốt quá trình học tập.

Cuối cùng, tôi xin gửi lời cảm ơn chân thành tới người yêu cùng bạn bè đã chia sẻ và giúp đỡ tôi trong suốt thời gian học tập trên giảng đường đại học và trong quá trình hoàn thành khóa luận. Tôi xin chân thành cảm ơn!

Hà Nội, ngày 4 tháng 12 năm 2023

Ký tên

Vũ Chí Dũng

LỜI CAM ĐOAN

Tôi cam đoan luận văn tốt nghiệp **“Phát triển kỹ thuật tư vấn công bằng và ứng dụng trong GIFT4Rec”** là công trình nghiên cứu của riêng tôi và được sự hướng dẫn của PGS. TS Hà Quang Thụy với ThS Nguyễn Trần Ngọc Linh chưa từng được nộp như một báo cáo khóa luận tại trường Đại học Công nghệ - Đại học Quốc gia Hà Nội hoặc bất kỳ trường đại học nào khác. Những gì tôi viết ra không sao chép từ các tài liệu, không sử dụng kết quả của người khác mà không trích dẫn cụ thể.

Tôi xin cam đoan những nội dung tôi trình bày trong khóa luận là do tôi tự phát triển, không sao chép mã nguồn của người khác. Nếu sai tôi xin chịu hoàn toàn trách nhiệm theo quy định của trường Đại học Công nghệ - Đại học Quốc gia Hà Nội.

Sinh viên thực hiện

Vũ Chí Dũng

MỤC LỤC

TÓM TẮT	i
LỜI CẢM ƠN	ii
LỜI CAM ĐOAN	iii
MỤC LỤC	iv
DANH MỤC HÌNH VẼ.....	vi
DANH MỤC BẢNG BIỂU	vii
BẢNG CHÚ GIẢI THUẬT NGỮ	viii
CHƯƠNG 1. GIỚI THIỆU BÀI TOÁN.....	1
1.1. Giới thiệu chung về hệ tư vấn	1
1.1.1. Hệ tư vấn dựa trên những người dùng giống nhau.....	2
1.1.2. Hệ tư vấn dựa trên những sản phẩm giống nhau	3
1.1.3. Hệ tư vấn dựa trên đồ thị	3
1.1.4. Hệ tư vấn theo phiên hay chuỗi tương tác tuần tự.....	4
1.2. Khởi động nguội trong hệ tư vấn.....	5
1.2.1. Sản phẩm khởi động nguội trong hệ tư vấn.....	6
1.2.2. Người dùng khởi động nguội trong hệ tư vấn	6
1.3. Công bằng trong hệ tư vấn	9
1.4. Phát biểu bài toán khóa luận.....	12
Phạm vi khóa luận	12
CHƯƠNG 2. NHỮNG VẤN ĐỀ LIÊN QUAN.....	13
2.1. Mô hình DropoutNet cho khởi động nguội	13
2.2. Sơ lược về mô hình GIFT4Rec cho khởi đầu nguội	17
2.3. Ý tưởng cải tiến của khóa luận	18
CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT	19
3.1. Mô hình GIFT4Rec	19
3.2. Mô hình cải tiến.....	22
CHƯƠNG 4. THỰC NGHIỆM VÀ ĐÁNH GIÁ.....	29
4.1. Phần cứng và phần mềm.....	29
Phần cứng	29
Phần mềm	29

4.2. Các kịch bản thực nghiệm	29
4.3. Dữ liệu thử nghiệm.....	29
4.4. Độ đo	30
4.5. Kết quả thực nghiệm.....	31
Kết luận	35
Tài liệu tham khảo.....	36

DANH MỤC HÌNH VẼ

- Hình 1.1: Minh họa cho lớp mô hình tư vấn dựa theo nội dung
- Hình 1.2: Minh họa cho lớp mô hình lọc cộng tác.
- Hình 1.3: Minh họa cách mô hình hóa hệ tư vấn thành đồ thị.
- Hình 1.4: Minh họa hệ tư vấn theo phiên hay chuỗi tương tác tuần tự.
- Hình 1.5: Mô tả giải pháp tư vấn người dùng khởi động nguội bằng chéo miền
- Hình 1.6: Mô tả tư vấn người dùng khởi động nguội sử dụng thông tin bổ trợ
- Hình 2.1: Kiến trúc chung của mô hình DropoutNet.
- Hình 3.1: Kiến trúc tổng quan của mô hình GIFT4Rec
- Hình 3.2: Kiến trúc tổng quan của mô hình cải tiến
- Hình 3.3: Mô-đun sinh trọng số.

DANH MỤC BẢNG BIỂU

Bảng 4.1: *Phân cứng sử dụng trong quá trình thực nghiệm*

Bảng 4.2: *So sánh hiệu năng mô hình đề xuất với các giải pháp đã có*

Bảng 4.3: *So sánh hiệu năng mô hình đề xuất với các biến thể của nó đã bỏ đi một thành phần cải tiến*

BẢNG CHÚ GIẢI THUẬT NGỮ

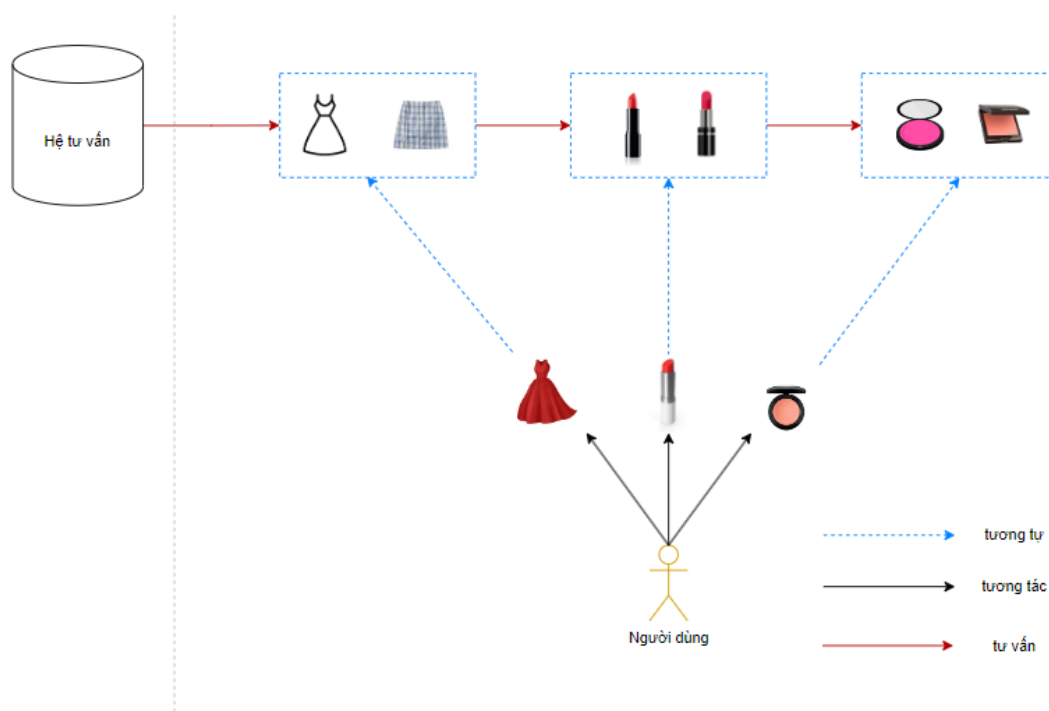
STT	Thuật ngữ	Ý nghĩa
1	Sản phẩm khởi động nguội	Là những sản phẩm mới hay chưa được bất cứ người dùng nào tương tác.
2	Người dùng khởi động nguội	Là những người dùng mới đăng ký vào hệ thống hay chưa có bất kỳ tương tác với sản phẩm nào.
3	Người dùng có tương tác	Là những người dùng đã có tương tác với sản phẩm trong hệ thống.
4	nhãn phản hồi ngầm	Nhãn sử dụng định nghĩa những tương tác của người dùng với sản phẩm nhưng thiếu đi chức năng đánh giá mức độ yêu thích như chấm điểm sản phẩm một cách chủ quan.

CHƯƠNG 1. GIỚI THIỆU BÀI TOÁN

1.1. Giới thiệu chung về hệ tư vấn

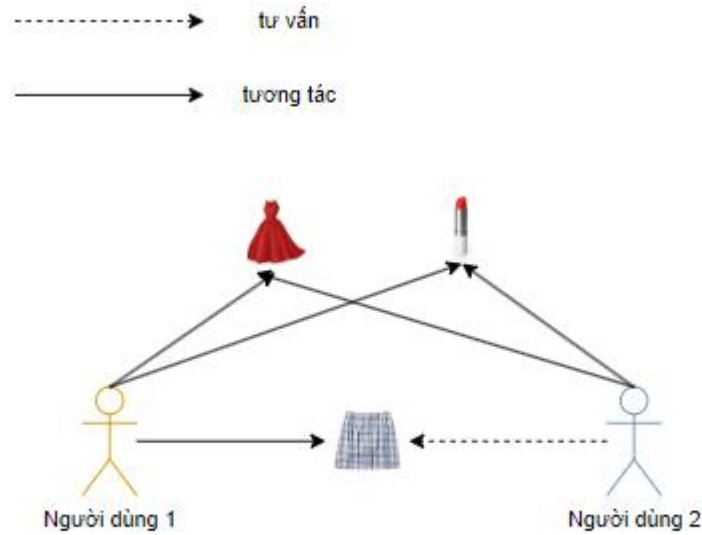
Trong thời đại công nghệ, với sự ra đời của rất nhiều sản phẩm công nghệ số cung cấp lượng lớn sản phẩm cho nhiều người dùng khác nhau, sự có mặt của một công cụ có thể giúp tìm kiếm những sản phẩm phù hợp cá nhân hóa với từng người dùng là điều vô cùng cần thiết. Điều này thúc đẩy sự hình thành của hệ tư vấn hay còn được gọi là hệ thống khuyến nghị với khả năng nắm bắt sở thích, hành vi của người dùng từ chuỗi tương tác của họ, từ đó đề xuất và đưa những sản phẩm cá nhân hóa. Hệ thống tư vấn đang được ứng dụng rộng rãi trong rất nhiều sản phẩm công nghệ lớn khác nhau, điển hình không thể không kể tới các nền tảng cung cấp phim với video trực tuyến như Netflix, Tik Tok hay các nền tảng thương mại điện tử như Shopee, Amazon và Alibaba. Các thuật toán tư vấn cổ điển bao gồm hai nhóm chính là lọc cộng tác và dựa trên nội dung.

Các thuật toán tư vấn dựa trên nội dung sẽ tư vấn những sản phẩm tương đồng với các sản phẩm người dùng có phản hồi. Cách tiếp cận này yêu cầu việc sắp xếp các sản phẩm vào từng nhóm hoặc đi tìm các đặc trưng của từng sản phẩm. Vậy nên, với những sản phẩm thiếu đi đặc trưng hay không có nhóm cụ thể thì sẽ không tận dụng được giải pháp này, đây là vấn đề cũng rất hay xảy ra trên thực tế với tình trạng thiếu dữ liệu hay thiếu giá trị trong dữ liệu.



Hình 1.1: Minh họa cho lớp mô hình tư vấn dựa theo nội dung

Thuật toán lọc cộng tác đánh giá xác định các sản phẩm phù hợp sở thích với mỗi người dùng từ những người dùng khác có hành vi hay mối quan tâm tương tự với họ. Dưới đây là minh họa cho thuật toán này.



Hình 1.2: Minh họa cho lớp mô hình lọc cộng tác.

Khóa luận này sẽ tập trung trình bày về các hệ tư vấn trong lớp mô hình lọc cộng tác.

1.1.1. Hệ tư vấn dựa trên những người dùng giống nhau

Xét tập người dùng U với số lượng người dùng là N_U , tập sản phẩm I với số lượng sản phẩm là N_I cùng với $X \in R^{U \times I}$ là ma trận thể hiện điểm số của từng người dùng cho từng sản phẩm đã quan sát được với $X_{u,i}$ là điểm số của người dùng u cho sản phẩm i . Để gợi ý các sản phẩm mới phù hợp với từng người dùng, thuật toán dự đoán điểm số của các sản phẩm chưa tương tác dựa trên những người dùng có hành vi tương tự với họ. Cụ thể như sau:

- Xét người dùng u , tìm ra những người dùng u' có hành vi tương tự với họ thông qua tính toán độ tương đồng giữa X_u và $X_{u'}$ sử dụng độ tương tự cosin.

$$\cos(X_u, X_{u'}) = \frac{X_u^T X_{u'}}{\|X_u\|_2 \cdot \|X_{u'}\|_2}$$

- Xét sản phẩm i' người dùng u chưa từng tương tác, $X_{u,i'}$ sẽ được dự đoán dựa theo tập $U_{u_similar_k}$ top k người dùng tương đồng nhất mà có tương tác với i' như sau:

$$X_{u,i'} = \frac{\sum_{u'} X_{u',i'} \times \cos(X_u, X_{u'})}{\sum_{u'} |\cos(X_u, X_{u'})|}$$

1.1.2. Hệ tư vấn dựa trên những sản phẩm giống nhau

Với các kí hiệu U , I và X tương tự như thuật toán nêu trên, định nghĩa X' là ma trận chuyển vị của X . Dự đoán điểm số của người dùng u chưa từng tương tác với sản phẩm i dựa vào những sản phẩm tương đồng nhất. Cụ thể như sau:

$$\cos(X'_i, X'_{i'}) = \frac{X'^T_i X'_{i'}}{\|X'_i\|_2 \cdot \|X'_{i'}\|_2}$$

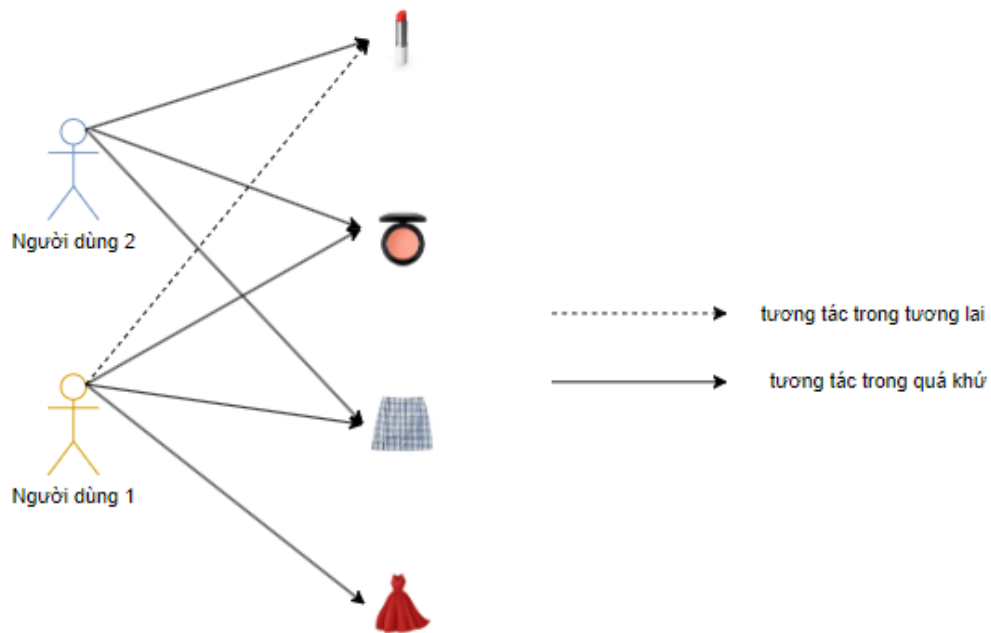
Tìm ra top k những sản phẩm i' được tương tác bởi người u tương đồng với i thông qua độ tương tự cosin giữa X'_i và $X'_{i'}$:

$$X'_{i,u} = \frac{\sum_{i'} X'_{i',u} \times \cos(X'_i, X'_{i'})}{\sum_{i'} |\cos(X'_i, X'_{i'})|}$$

1.1.3. Hệ tư vấn dựa trên đồ thị

Thuật toán này hiện tại được triển khai và xây dựng thành công bởi vô cùng nhiều công ty lớn. Trong đó có sàn thương mại điện tử Alibaba, một trong những sàn thương mại điện tử lớn nhất Trung Quốc. Việc triển khai thuật toán này thành một hệ thống tư vấn cho Viettel Mall là hoàn toàn khả thi. Cụ thể hơn, hiện nay Pinterest đã xây dựng và triển khai thành công hệ tư vấn dựa trên đồ thị với những cải tiến về mặt tối ưu thời gian chạy cũng như khiến mô hình hội tụ nhanh hơn trong bài toán tư vấn sản phẩm tương tự. Nguồn là những cặp sản phẩm đã được gán nhãn có liên kết hay không cùng thông tin có sẵn của chúng. Hệ tư vấn có thể mô hình hóa dưới dạng đồ thị hai phía.

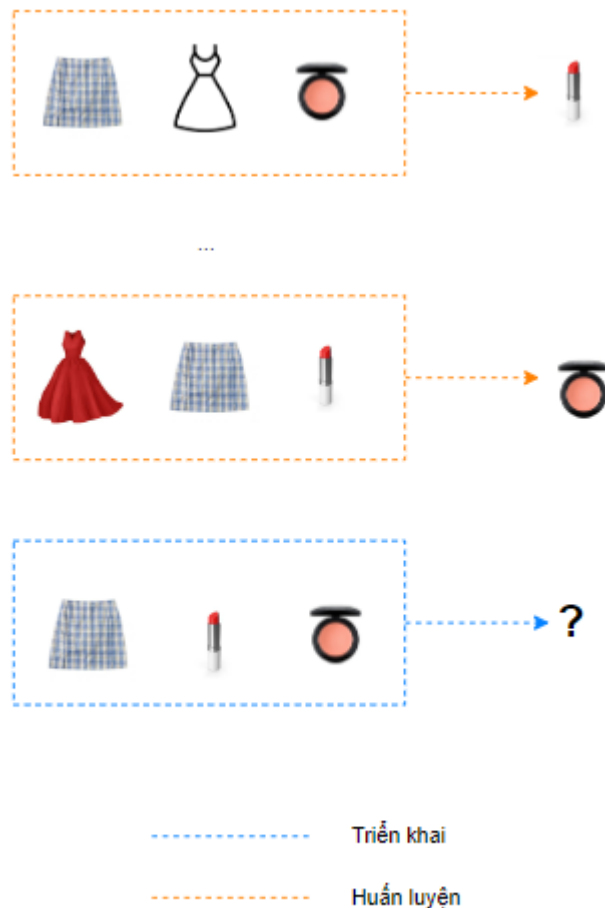
- 1 đồ thị gồm hai loại nút là sản phẩm và người dùng. Cạnh kết nối giữa các loại nút thường là tương tác của người dùng với sản phẩm được gán nhãn hành vi.
- Đầu vào bao gồm thông tin có sẵn của người dùng và sản phẩm cùng tương tác quan sát được
- Hệ tư vấn được đưa về bài toán dự đoán liên kết hay dự đoán tương tác mới giữa người dùng và sản phẩm dựa trên các tương tác trong quá khứ.



Hình 1.3: Minh họa cách thức mô hình hóa hệ tư vấn thành đồ thị

1.1.4. Hệ tư vấn theo phiên hay chuỗi tương tác tuần tự

Sau đây khóa luận sẽ trình bày về định nghĩa của nhóm mô hình này được nêu ra tại bài nghiên cứu đề xuất bởi Shoujin Wang và cộng sự tại [1]. Phương pháp này lấy tương tác tuần tự với sản phẩm của người dùng hay từ phiên làm đầu vào huấn luyện. Phiên có thể hiểu là một tập hợp các tương tác với sản phẩm liên tục, có thể được sắp xếp hoặc không có thứ tự. Một chuỗi tương tác tuần tự là một danh sách các sản phẩm hay đại diện của sản phẩm được tương tác và sắp xếp theo thứ tự thời gian rõ ràng. Hệ tư vấn theo phiên dự đoán sản phẩm tiếp theo được tương tác dựa trên các tương tác quá khứ nhất định trong một phiên hoặc dự đoán phiên trong tương, chủ yếu phụ thuộc vào sự phụ thuộc trong hoặc giữa các phiên. Bên cạnh đó, hệ tư vấn theo chuỗi tương tác tuần tự dự đoán các sản phẩm tiếp theo của chuỗi. Nhóm phương pháp này được mô tả như bên dưới.



Hình 1.4: Minh họa hệ tư vấn theo phiên hay chuỗi tương tác tuần tự

1.2. Khởi động nguội trong hệ tư vấn

Khởi động nguội là một thử thách lớn trong hệ tư vấn với viễn cảnh hệ tư vấn cần đưa ra đánh giá phù hợp cho người dùng mới hoặc sản phẩm mới hoàn toàn không có hoặc rất ít tương tác mà quan sát được. Vấn đề này trở nên vô cùng cần thiết do sự gia tăng nhanh về lượng người dùng và sản phẩm trong các nền tảng công nghệ số lớn hiện nay.

Khởi động nguội có thể chia làm hai nhóm chính:

- Thứ nhất là người dùng khởi động nguội, vấn đề này xảy ra khi người dùng lần đầu sử dụng dịch vụ nên chưa có lịch sử tương tác của họ để hệ tư vấn nắm bắt hành vi và cung cấp được những sản phẩm phù hợp.
- Thứ hai là vấn đề sản phẩm khởi động nguội, vấn đề sản phẩm khởi động nguội xảy ra khi các sản phẩm mới được cung cấp thêm trong hệ thống không có hoặc có rất ít tương tác.

Tất cả những vấn đề trên gây ra khó khăn đối với các hệ tư vấn lọc cộng tác do chúng dựa vào sự tương tác để đề xuất sản phẩm phù hợp.

1.2.1. Sản phẩm khởi động nguội trong hệ tư vấn

Trong hầu hết các sản phẩm thực tế, việc đề xuất các sản phẩm mới không có bất kỳ phản hồi nào từ người dùng là vô cùng khó khăn và thách thức. Trong nhiều năm qua, nhiều phương pháp đã được đề xuất bao gồm các thuật toán phi tham số huấn luyện được đề xuất bởi Suvash và cộng sự [2] bên cạnh các thuật toán học máy và học sâu. Các phương pháp này có thể được phân loại thành hai loại là huấn luyện riêng biệt và huấn luyện đồng thời. Trước tiên là các phương pháp huấn luyện riêng biệt [3, 4, 5, 6, 7] với mục đích huấn luyện riêng hai mô hình:

- Một mô hình tư vấn lọc cộng tác để học biểu diễn của các sản phẩm cũ đã có tương tác.
- Một mô hình tư vấn dựa trên nội dung để học cách ánh xạ từ nội dung một sản phẩm sang biểu diễn từ tương tác của sản phẩm thu được từ mô hình tư vấn lọc cộng tác đã được đề cập bên trên.

Trong quá trình triển khai, biểu diễn của sản phẩm cũ sẽ là đầu ra của mô hình tư vấn lọc cộng tác còn biểu diễn của sản phẩm mới sẽ được ánh xạ từ nội dung của chúng qua mô hình tư vấn dựa trên nội dung. Một ví dụ điển hình là mô hình DeepMusic [7], sử dụng mạng nơ-ron nhiều lớp để học cách ánh xạ nội dung sản phẩm sang biểu diễn được học bằng mô hình phân rã ma trận được huấn luyện trước.

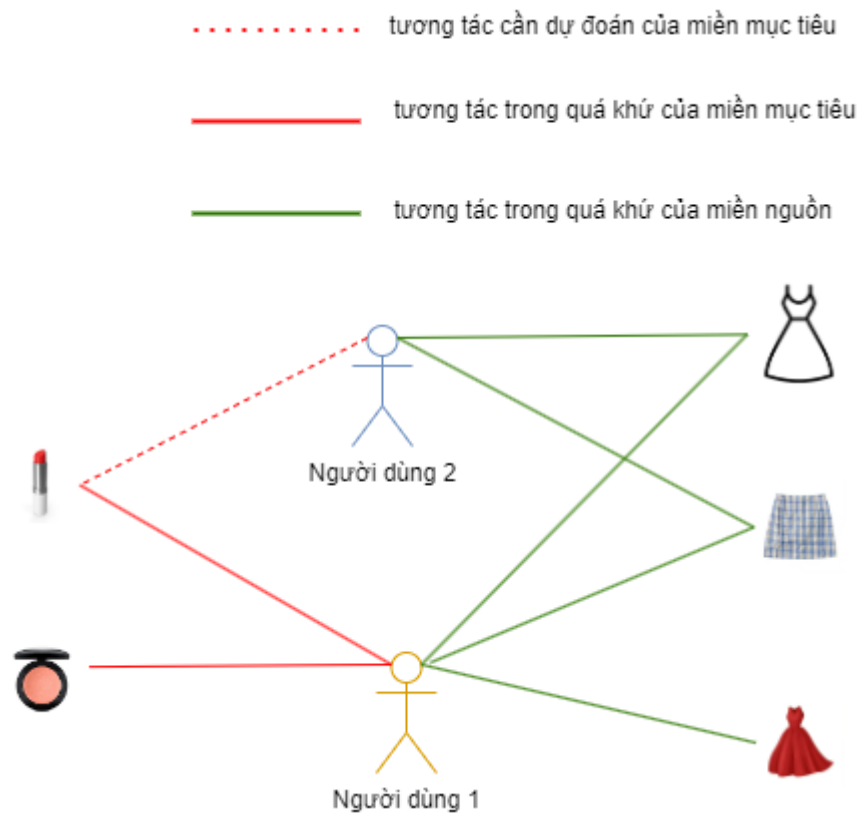
Phương pháp huấn luyện đồng thời [8, 9, 10] kết hợp mô hình tư vấn lọc cộng tác và mô hình tư vấn dựa trên nội dung với nhau và huấn luyện cả hai đồng thời thông qua một quá trình lan truyền ngược duy nhất. Một ví dụ điển hình là thuật toán DropoutNet [10] có mô hình tư vấn dựa trên nội dung sử dụng mạng nơ-ron nhiều lớp để sẽ được ánh xạ từ nội dung của chúng thành biểu diễn của sản phẩm, sau đó là mô hình tư vấn lọc cộng tác nhằm mục tiêu cung cấp tư vấn.

1.2.2. Người dùng khởi động nguội trong hệ tư vấn

Hiện nay, theo như hiểu biết của khóa luận, có hai nhóm phương pháp chính giải quyết vấn đề người dùng khởi động nguội.

Nhóm thứ nhất bao gồm các nhóm mô hình lấy nền tảng từ các giải pháp xử lý bài toán miền chéo. Bài toán miền chéo là bài toán tận dụng nguồn tương tác từ một miền chéo để bù đắp cho sự thừa thớt dữ liệu tương tác trong miền sản phẩm tiêu và nâng cao hiệu năng của hệ tư vấn. Đối với bài toán người dùng khởi động nguội trong hệ tư vấn, các nhóm mô hình trong nhóm thứ nhất tận dụng nguồn tri thức và tương tác từ miền chéo quan sát được của người dùng khởi động nguội trong miền sản phẩm tiêu để gia tăng hiệu năng tư vấn cho họ [11, 12]. Những bộ lọc cộng tác này dựa trên hoạt động bị ảnh hưởng nghiêm trọng bởi vấn đề thừa thớt dữ liệu khi quan sát toàn bộ các miền chéo nhau. Tuy nhiên, nhóm giải pháp này gặp một hạn chế lớn khi mà trên thực tế không phải khi nào cũng có miền nguồn để lấy tương tác bù đắp cho sự thừa thớt dữ

liệu của miền mục tiêu. Chính vì vậy, trong thực tế và môi trường làm sản phẩm, người ta hay trích xuất thông tin hành vi từ những thông tin có sẵn của người dùng cung cấp khi đăng ký sử dụng dịch vụ như giới tính, độ tuổi và vị trí địa lý tương đối. Chi tiết về giải pháp này sẽ được mô tả bên dưới. Nhóm giải pháp thứ nhất có thể mô tả như trong hình vẽ 1.5 như sau:



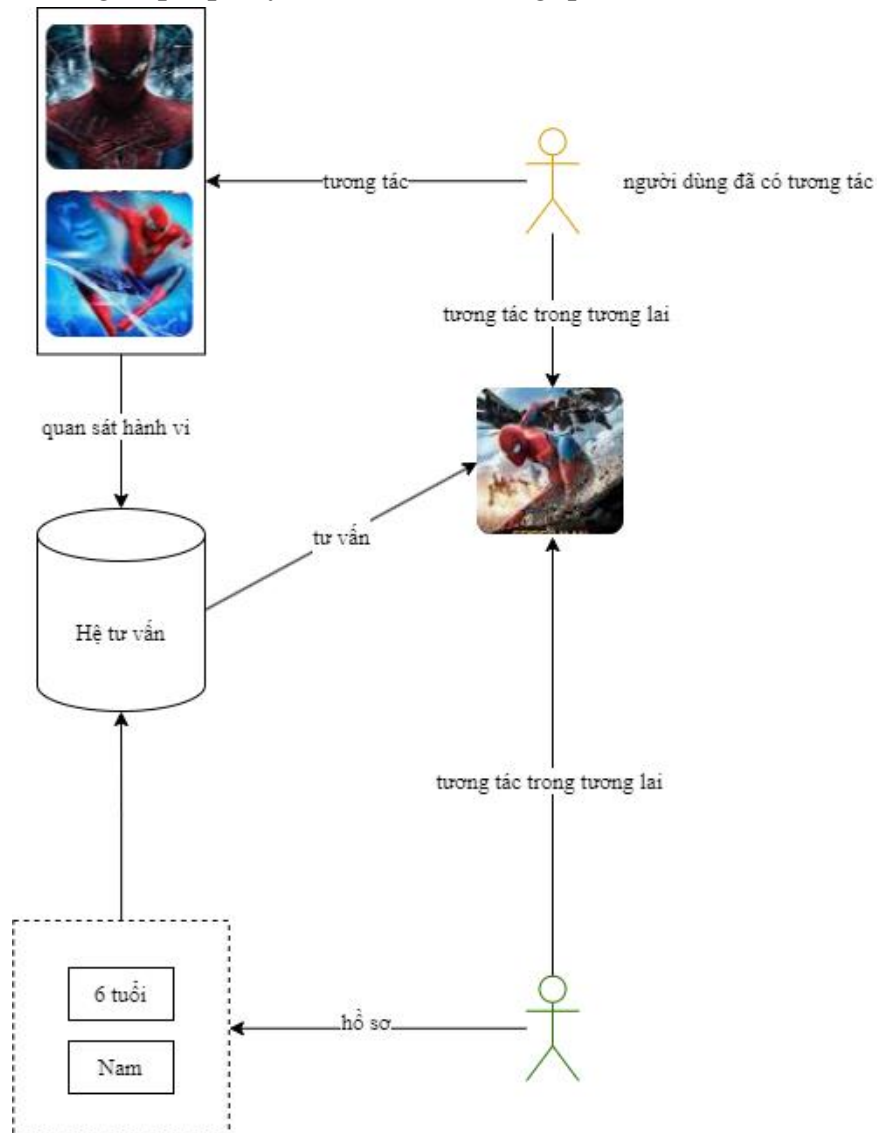
Hình 1.5: Mô tả giải pháp tư vấn người dùng khởi động nguội bằng chéo miền

Nhóm Thứ hai bao gồm các giải pháp hợp nhất các trường thông tin bổ trợ. Đối với bài toán người dùng khởi động nguội, mô hình tận dụng các trường thông tin bổ trợ được người dùng cung cấp lúc đăng ký dùng dịch vụ như giới tính, độ tuổi và vị trí địa lý tương đối; kết hợp hoặc song song với thông tin hành vi trích xuất từ tương tác với sản phẩm mà quan sát được với mục tiêu nâng cao hiệu năng của hệ tư vấn. Giải pháp xử lý này rất phù hợp cho bài toán người dùng khởi động nguội, giúp hệ tư vấn có thể phán đoán được những hành vi nhất định của người dùng mà không cần đến tương tác. Lấy ví dụ như sau:

- Những người dùng có độ tuổi thiếu nhi sử dụng nền tảng công nghệ số cung cấp phim, video thì có thể sẽ thích hoạt hình.
- Những người giới tính nữ khi mua hàng online sẽ có xu hướng thích những món hàng dành cho đúng giới tính như váy vóc, mỹ phẩm nhiều hơn.

Những phương pháp đầu tiên tập trung vào việc sử dụng thông tin bổ trợ như hồ sơ người dùng và đặc tính của sản phẩm để đánh giá sở thích của người dùng khởi động

ngươi. Ngoài ra, nhiều mô hình kết hợp tích hợp thông tin bên lề vào lọc cộng tác cũng được đề xuất như [13, 14]. Ví dụ, hợp tác hồi quy chủ đề [15] kết hợp mô hình chủ đề xác suất [16] với mô hình phân rã ma trận để nâng cao khả năng dự đoán ngoài ma trận của mô hình. Các giải pháp này có thể mô tả thông qua hình vẽ 1.6 dưới đây:



Hình 1.6: Mô tả tư vấn người dùng khởi động người sử dụng thông tin bổ trợ

Các giải pháp hợp nhất thêm các thông tin bổ trợ của người dùng hiện nay đã giải quyết vấn đề người dùng khởi động người nhưng đa phần đều vấp phải một nhược điểm lớn. Việc kết hợp thông tin bổ trợ của người dùng một cách không kiểm soát có thể gây nhiễu và khiến mô hình trở nên quá khớp, tuy giải quyết được vấn đề người dùng khởi động người nhưng lại giảm hiệu năng tư vấn trên những người dùng cũ đã có tương tác và khiến hệ tư vấn trở nên thiếu công bằng. Cho đến hiện tại, theo hiểu biết của tôi mới có hai giải pháp được nêu dưới đây chú ý và giải quyết sự thiếu công bằng này là mô

hình GIFT4Rec [17] và DropoutNet [10]. Đây cũng chính là vấn đề đặt ra làm động lực của khóa luận này ra đời.

1.3. Công bằng trong hệ tư vấn

Công bằng là một khái niệm bắt nguồn từ lĩnh vực xã hội học, kinh tế và luật. Trong ngữ cảnh của hệ tư vấn, sự công bằng đòi hỏi người giới thiệu đồ hệ thống đối xử bình đẳng với tất cả người dùng và sản phẩm. Ví dụ, trong bài toán phê duyệt khoản vay sử dụng hệ tư vấn, hệ thống này được đánh giá là công bằng nếu việc chấp nhận khoản vay không ảnh hưởng bởi hồ sơ của họ như giới tính hay chủng tộc. Định nghĩa về thiên vị theo như bài nghiên cứu của Di Jin và cộng sự [18] được trình bày như sau:

Vòng đời của hệ tư vấn bao gồm ba giai đoạn là giai đoạn thu thập dữ liệu, huấn luyện mô hình và vòng lặp phản hồi. Sự bất công trong hệ tư vấn có thể xuất hiện trong toàn bộ vòng đời của nó, bao gồm cả giai đoạn thu thập dữ liệu lẫn giai đoạn huấn luyện mô hình và giai đoạn vòng lặp phản hồi. Dựa theo từng giai đoạn trong vòng đời hệ tư vấn mà các loại thiên vị xuất hiện, Di Jin và cộng sự đã chia chúng thành ba nhóm chính là thiên vị về mặt dữ liệu, mô hình và phản hồi. Trước tiên là dạng thiên vị về mặt dữ liệu. Hiện nay, hệ tư vấn luôn được huấn luyện trên các tập dữ liệu tương tác và hành vi lớn, thường chứa những thông tin như tương tác người dùng và hồ sơ của họ. Tuy nhiên, tập dữ liệu huấn luyện có khả năng chứa đựng những thiên vị trong đó. Những thiên vị này có thể dẫn đến việc tư vấn sản phẩm không công bằng, gây ra những hậu quả không mong muốn hoặc thậm chí tai hại, nghiêm trọng đối với đời sống con người và xã hội. Trước tiên, khóa luận sẽ trình bày về thiên vị trong mặt dữ liệu:

- Thứ nhất là thiên vị về mặt người dùng. Thiên vị phổ biến trong nhóm này là thiên vị về mặt thuộc tính. Trong hệ tư vấn một số thuộc tính nhạy cảm của người dùng như tuổi, vị trí địa lý, giới tính và nghề nghiệp là những nguồn thông tin quan trọng để nắm bắt được sở thích của người dùng. Tuy nhiên, đôi khi thuộc tính của người dùng (ví dụ: tuổi, giới tính) có thể gây ra kết quả tư vấn thiên vị.
- Tiếp theo là thiên vị đến từ sự giới hạn về lượng sản phẩm có thể phơi bày ra trước mắt người dùng, điều này dẫn đến một hệ tư vấn thiếu công bằng. Lấy ví dụ với một hệ tư vấn âm nhạc, tác phẩm từ một số nghệ sĩ nổi tiếng có thể thu hút phần lớn lưu lượng truy cập, dẫn đến làm lu mờ các nghệ sĩ khác và xảy ra tình trạng thiếu công bằng.
- Cuối cùng chính là vấn đề thiên vị khởi động nguội, cụ thể về dạng thiếu công bằng này đã được trình bày cụ thể ở chương 1.2

Tiếp theo là dạng thiên vị trong mô hình. Các nhóm mô hình thuộc hệ tư vấn cố gắng nắm bắt sở thích của người dùng từ dữ liệu tương tác trong quá khứ để dự đoán xác suất người dùng chọn những sản phẩm mới. Các mô hình thiết kế chưa tối ưu có thể làm tăng thêm thiên vị trong dữ liệu đầu vào hay thiên vị về thứ tự sắp xếp các sản phẩm tư vấn hiển thị cho người dùng. Một số hàm mất mát có thể làm cho sự thiếu công bằng trong các hệ tư vấn trầm trọng hơn và ảnh hưởng đến đầu ra là khả năng người dùng tương tác sản phẩm, từ đó gây tác động không tốt đến thứ tự xếp hạng các sản phẩm hệ tư vấn đề xuất. Wan và cộng sự [19] đã chứng minh rằng các hàm mất mát thường sử dụng trong hệ tư vấn rất nhạy cảm với các sản phẩm phổ biến và khiến mô hình luôn

đánh giá điểm số cao hơn so với những sản phẩm không phổ biến, điều này làm tăng thiên vị trong quá trình huấn luyện hệ tư vấn.

Cuối cùng là thiên vị trong phản hồi. Một vòng lặp phản hồi tồn tại trong mọi hệ tư vấn, cung cấp kết quả tư vấn từ mô hình để người dùng lựa chọn. Trong quá trình này, hành vi mới của người dùng được hệ tư vấn sử dụng để huấn luyện thêm cho mô hình và trả ra một danh sách sản phẩm tư vấn tiếp theo. Quá trình này có thể tạo ra cũng như làm sự thiên vị trầm trọng hơn theo thời gian, dẫn đến sự suy thoái của tính công bằng trong hệ tư vấn. Thiên vị về mức độ phổ biến là một dạng thiên vị trong phản hồi điển hình, loại thiên vị này đề cập đến vấn đề một số sản phẩm phổ biến thường xuyên được tư vấn và hầu hết phần còn lại bị bỏ qua. Người dùng sử dụng những sản phẩm được tư vấn này và phản hồi của họ lại được cập nhật vào mô hình của hệ tư vấn. Ngày qua ngày, hệ tư vấn sẽ chỉ đề xuất các sản phẩm phổ biến cho người dùng, liên tục thu thập phản hồi của họ về các sản phẩm phổ biến rồi lại thêm chúng vào huấn luyện, khiến cho phân bố dữ liệu trở nên mất cân bằng hơn cũng như mô hình dần chỉ tư vấn các sản phẩm phổ biến. Sự tồn tại của loại thiên vị này sẽ gây ảnh hưởng xấu tới biểu diễn của người dùng mà hệ tư vấn học được và gây khó khăn trong việc nắm bắt sở thích của họ.

Tiếp theo, khóa luận sẽ trình bày về định nghĩa các biểu hiện của một hệ tư vấn công bằng [18] dưới các góc nhìn chính dưới đây:

- Biểu hiện thứ nhất của một hệ tư vấn công bằng là công bằng cá nhân. Với hai người dùng tương tự nhau, hệ tư vấn có biểu hiện thứ nhất này sẽ luôn đưa ra tư vấn mang chất lượng tương đương. Lấy ví dụ về một hệ tư vấn bên lĩnh vực chăm sóc sức khỏe, hai bệnh nhân có bệnh lý giống nhau sẽ nhận được tư vấn tương đương nhau về mặt chất lượng bất kể tình trạng giàu nghèo.
- Biểu hiện thứ hai của một hệ tư vấn công bằng là công bằng giữa các nhóm người dùng hay nói cách khác phải đưa ra tư vấn có chất lượng tương đương giữa các nhóm người dùng khác nhau.
- Biểu hiện tiếp theo được thể hiện qua sự công bằng trong xếp hạng các sản phẩm sử dụng đầu ra của mô hình hệ tư vấn. Sự công bằng trong việc xếp hạng thường được phản ánh trong sự công bằng về mặt thống kê hoặc sự công bằng về cơ hội. Nói chung, sự công bằng trong xếp hạng yêu cầu các sản phẩm hoặc nhóm sản phẩm tương tự phải nhận được khả năng hiển thị tương đương nhau, cụ thể hơn là xuất hiện ở các vị trí tương tự trong đầu ra xếp hạng của mô hình hệ tư vấn.

Hiện tại, có 3 nhóm giải pháp chính can thiệp vào các giai đoạn khác nhau xây dựng một hệ tư vấn với mục tiêu nâng cao tính công bằng. Trước tiên là các giải pháp tiền xử lý cho hệ tư vấn công bằng. Các giải pháp trong nhóm này nâng cao tính công bằng và giảm thiểu thiên vị trong tập dữ liệu huấn luyện của mô hình. Những thiên vị này có thể được khuếch đại trong suốt vòng đời của một hệ tư vấn, dẫn đến sự suy thoái dần dần của tính công bằng trong hệ tư vấn. Trong nhóm giải pháp này tồn tại 3 nhóm giải pháp con được mô tả bên dưới như sau:

- Giải pháp đầu tiên là gán nhãn lại dữ liệu huấn luyện. Giải pháp này sẽ thay đổi nhãn của tập huấn luyện cho mô hình nhằm giảm thiểu sự sai lệch trong đầu vào. Một số hệ tư vấn sử dụng nhãn phản hồi ngầm dạng nhị phân như "nhấp chuột" hay "không nhấp chuột" vào sản phẩm, điều này có thể gây nhiễu do người dùng có thể bấm nhầm hay tiêu chí này chưa đủ để đánh giá phản hồi tích cực hay tiêu cực, từ đó gây ảnh hưởng không tốt đến hiệu năng của hệ tư vấn. Wang và cộng

sự [20] đã thiết kế một phương pháp học bán quan sát tự động gán nhãn lại cho những thành phần dữ liệu gây nhiễu trong phản hồi ngầm. Phương pháp này tự động sinh ra các nhãn giả định hình sở thích của người dùng để giảm thiểu nhiễu trong cả hai thể loại phản hồi quan sát được và phản hồi ngầm.

- Giải pháp tiếp theo là tái phân phối lại tập dữ liệu huấn luyện. Mất cân bằng trong phân phối dữ liệu có thể làm hiệu năng của hệ tư vấn giảm trầm trọng. Montanari và cộng sự [21] đề xuất một thuật toán chọn mẫu dữ liệu để đảm bảo rằng việc lấy mẫu đồng đều và không bị ảnh hưởng bởi phân bố. Họ chọn ngẫu nhiên một tỷ lệ phần trăm người dùng nhất định và xóa thông tin tương tác với sản phẩm của họ.
- Giải pháp cuối cùng là chỉnh sửa lại dữ liệu huấn luyện. Ý tưởng chính là sinh thêm hay sửa đổi dữ liệu sai lệch để giảm thiên vị. Lấy ví dụ, một số hệ tư vấn có thể dùng đến dữ liệu văn bản như hệ tư vấn việc làm hay hệ tư vấn tin tức, trong đó có thể bị thiếu hoặc có sai sót và gây ra thiên vị. Để giải quyết vấn đề này, Wang và cộng sự [22] đã sử dụng các kỹ thuật tiền xử lý trong lĩnh vực ngôn ngữ tự nhiên để sửa đổi dữ liệu huấn luyện nâng cao hiệu năng tư vấn.

Tiếp theo là các giải pháp can thiệp vào quá trình huấn luyện hay triển khai với mục tiêu học được hệ tư vấn công bằng. Nhóm phương pháp này được phân thành các nhóm chính bao gồm phương pháp dựa trên điều chuẩn, phương pháp dựa trên suy luận thông thường, phương pháp dựa trên học đối kháng, phương pháp dựa trên học tăng cường và phương pháp xếp hạng.

Cuối cùng là những giải pháp gia tăng tính công bằng của hệ tư vấn bằng cách hậu xử lý, sử dụng các phương pháp hậu xử lý nhằm mục đích xếp hạng lại kết quả tư vấn là đầu ra từ các mô hình một cách thủ công hoặc dựa trên thuật toán. Khóa luận sẽ tập trung giới thiệu và trình bày thuật toán xếp hạng lại giúp tăng cường tính công bằng trong hệ tư vấn cùng hai nhóm chính là không sử dụng tham số và có sử dụng tham số.

Trước tiên là phương pháp xếp hạng lại không sử dụng tham số huấn luyện. Hay các thuật toán không cần học tập để tiến hành xếp hạng lại liên quan đến tính công bằng. Phương pháp heuristic được sử dụng để đạt sự công bằng trong hệ tư vấn có thể coi các chiến lược phi tham số, coi quá trình xếp hạng lại như một bài toán quy hoạch tuyến tính. Từ các ràng buộc công bằng cho hệ tư vấn được định sẵn, thuật toán sẽ tìm kiếm heuristic ra bộ danh sách xếp hạng lại có kết quả tối ưu nhất.

Tiếp theo là giải pháp xếp hạng lại có sử dụng tham số để huấn luyện. Zhu và cộng sự [23] huấn luyện một mạng bộ tự mã hóa để nâng cao tính công bằng và đảm bảo kết quả được xếp hạng lại có cùng phân phối với ban đầu.

Cuối cùng, khóa luận sẽ trình bày về công thức đánh giá tính công bằng của một hệ tư vấn. Một công thức phổ biến thường xuyên được sử dụng biểu thị sự công bằng giữa các nhóm người dùng khác nhau được trình bày trong bài nghiên cứu đề xuất bởi Yunqi Li và cộng sự [24]. GF là viết tắt của Group Fairness cùng với A_1 và A_2 là hai nhóm người dùng khác nhau. $R(.)$ Là công thức đánh giá hiệu năng bài toán đặt ra. Độ đo nhận giá trị càng nhỏ thì hệ tư vấn càng công bằng.

$$GF = \left| \frac{1}{|A_1|} \sum_{u_1 \in A_1} R(u_1) - \frac{1}{|A_2|} \sum_{u_2 \in A_2} R(u_2) \right|$$

1.4. Phát biểu bài toán khóa luận

Bài toán người dùng khởi động nguội đang là vấn đề thời sự trong hệ tư vấn do sự gia tăng chóng mặt về lượng người dùng của các nền tảng công nghệ số lớn đang ứng dụng hệ tư vấn. Để giải quyết vấn đề này, các nhóm mô hình hiện tại kết hợp thông tin hỗ trợ hoặc sử dụng giải pháp dựa trên chéo miền. Trên thực tế không phải lúc nào cũng có miền chéo trong bài toán để sử dụng giải pháp dựa trên chéo miền. Chính vì vậy, bài toán của khóa luận tập trung vào nhóm giải pháp kết hợp thêm các thông tin hỗ trợ. Tuy nhiên phần lớn các giải pháp trong nhóm này chỉ giải quyết bài toán người dùng khởi động nguội mà lại bỏ qua vấn đề duy trì hiệu năng tư vấn trên người dùng cũ, gây nên tình trạng thiếu công bằng giữa hai nhóm người dùng đã có tương tác và khởi động nguội.

Lấy vấn đề này làm động lực, bài toán khóa luận sẽ tập trung vào giải quyết vấn đề người dùng khởi động nguội nhưng vẫn duy trì được hiệu năng tư vấn trên tập người dùng đã có tương tác, hướng đến một hệ tư vấn công bằng.

Phạm vi khóa luận

Khóa luận này tập trung nghiên cứu và phát triển mô hình GIFT4Rec [17] kết hợp thông tin hỗ trợ của người dùng sử dụng trọng số cùng phương pháp huấn luyện đa tác vụ để tăng hiệu năng tư vấn. Bên cạnh đó, khóa luận cũng nghiên cứu và vận dụng cách thức huấn luyện trong mô hình DropoutNet [10] được đề xuất bởi Maksims Volkovs với Guangwei Yu vào mô hình đề xuất trong khóa luận.

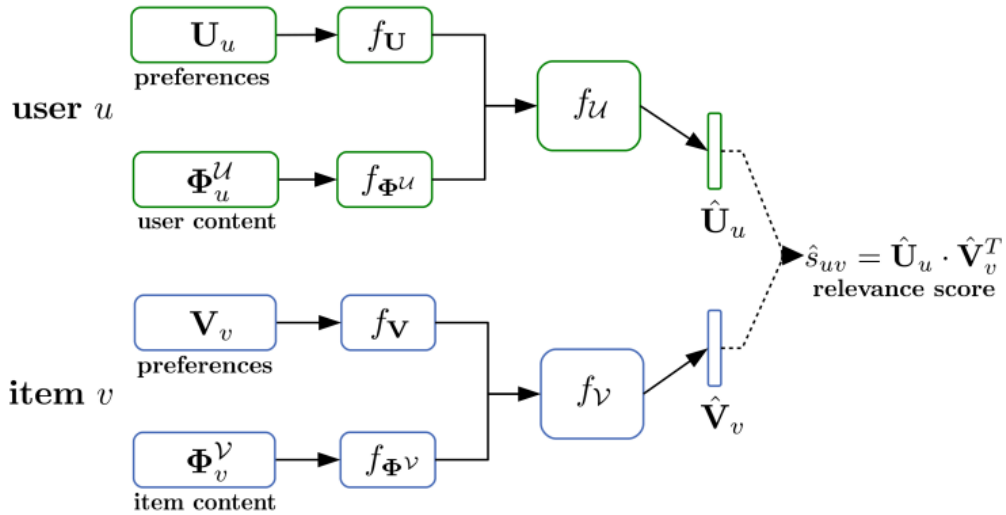
Kết luận

Trong chương 1, khóa luận đã giới thiệu về hệ tư vấn cùng các vấn đề thời sự là khởi động nguội và công bằng. Khóa luận đã chỉ ra rằng đa phần các phương pháp dùng thêm thông tin hỗ trợ giải quyết bài toán người dùng khởi động nguội mà lại bỏ qua vấn đề duy trì hiệu năng tư vấn trên người dùng cũ. Đây chính là động lực để khóa luận phát triển mô hình GIFT4Rec để giải quyết bài toán khóa luận. Trong chương tiếp theo, khóa luận sẽ trình bày về các nhóm giải pháp giải quyết được vấn đề người dùng khởi động nguội mà vẫn duy trì được hiệu năng tư vấn trên người dùng có tương tác và nêu ra ý tưởng cải tiến của khóa luận.

CHƯƠNG 2. NHỮNG VẤN ĐỀ LIÊN QUAN

2.1. Mô hình DropoutNet cho khởi động nguội

Mô hình DropoutNet được đề xuất trong bài “DropoutNet: Addressing Cold Start in Recommender Systems” của tác giả Maksims Volkovs với Guangwei Yu, được công bố tại hội nghị NIPS vào năm 2017 [10]. Dưới đây là hình vẽ kiến trúc chung của mô hình được khóa luận trích dẫn từ bài nghiên cứu gốc.



Hình 2.1: Kiến trúc chung của mô hình DropoutNet

Đối với mỗi người dùng u , biểu diễn từ tương tác là U_u và các thông tin bổ trợ Φ_u^U lần lượt được ánh xạ qua các hàm f_U và f_{Φ^U} rồi nối lại với nhau xong đưa qua mạng tinh chỉnh f_u thu được biểu diễn chính thức là \hat{U}_u . Các thành phần của sản phẩm v tương tự với f_V , f_{Φ^V} và f_v để tạo ra \hat{V}_v .

Mô hình sử dụng các thành phần riêng biệt ánh xạ biểu diễn từ tương tác và thông tin bổ trợ với 2 lý do chính:

- Xử lý dễ dàng hơn các đầu vào của dữ liệu có cấu trúc phức tạp trong thông tin bổ trợ chẳng hạn như hình ảnh hay văn bản chưa thể kết hợp trực tiếp ngay với biểu diễn tương tác qua phép nối.
- Cho phép sử dụng các mô hình đã được tiền huấn luyện cho f_{Φ^U} và f_{Φ^V} . Thời gian tiêu tốn để huấn luyện mô hình có thể được giảm thiểu đáng kể chỉ bằng việc cập nhật một vài lớp cuối cùng của mô hình được tiền huấn luyện.

Nếu phân phối của thông tin bổ trợ và biểu diễn từ tương tác được đánh giá là có phân phối giống nhau và tương thích thì mô hình sẽ nối trực tiếp chúng với nhau từ đầu vào và đưa trực tiếp vào mạng tinh chỉnh. Ví dụ với người dùng u , nếu trong viễn cảnh

như bên trên đề cập tới thì U_u và Φ_u^U sẽ được nối với nhau và trực tiếp được ánh xạ qua f_U để thu về U_u^\wedge .

Tất cả các thành phần được tối ưu đồng thời trong quá trình lan truyền ngược và sau đó được cố định trong quá trình triển khai. Khi mô hình đã được huấn luyện, chúng cố định nó và ánh xạ biểu diễn của người dùng U tới biểu diễn mới là U^\wedge và biểu diễn của sản phẩm V tới biểu diễn mới là V^\wedge qua mô hình. Tất cả công việc sau đó được thực hiện sử dụng hai biểu diễn mới là U^\wedge và V^\wedge cùng khả năng người dùng u tương tác sản phẩm v mô hình dự đoán sẽ được tính toán bởi công thức:

$$s_{u,v}^\wedge = U_u^\wedge (V_v^\wedge)^T$$

Trong quá trình huấn luyện, mô hình cố gắng giải quyết được vấn đề khởi động nguội nhưng vẫn đảm bảo hiệu năng tư vấn trên những người dùng và sản phẩm đã có tương tác. Các mô hình hiện tại xử lý vấn đề khởi động nguội bằng cách bổ sung thêm các thông tin hỗ trợ cũng như huấn luyện mô hình để tư vấn dựa trên thông tin hỗ trợ thuần túy dự phòng trong trường hợp thiếu đi biểu diễn tương tác. Tuy nhiên, giải pháp này làm việc học trở nên phức tạp hơn khi phải tối ưu các tác vụ khác nhau. Hơn nữa, thông tin hỗ trợ không phải lúc nào cũng có ích trong việc đánh giá hành vi của người dùng và có thể gây nhiễu cũng như khiến mô hình trở nên quá khớp. Lấy những nhược điểm còn tồn đọng được nêu ra bên trên của các giải pháp đã có làm động lực, tác giả bài báo tiếp cận theo cách mới mượn ý tưởng từ bộ mã hóa tự động khử nhiễu và huấn luyện mô hình để tái cấu trúc biểu diễn thực một cách chính xác ngay cả khi khuyết thiếu thông tin. Nhằm đạt được mục tiêu, tác giả bài báo đã đề xuất tối ưu một hàm mất mát như sau:

$$O = \sum_{u,v} (U_u V_v^T - f(U_u, \Phi_u^U) f(V_v, \Phi_v^V)^T)^2 = \sum_{u,v} [(U_u V_v^T - U_u^\wedge (V_v^\wedge)^T)]^2$$

Hàm mất mát này giảm thiểu lỗi giữa khả năng tương tác được dự đoán từ mô hình biểu diễn tương tác được tiền huấn luyện và mô hình kết hợp thêm thông tin hỗ trợ. Điều này giúp mô hình ngay cả khi kết hợp thêm thông tin hỗ trợ sẽ duy trì được độ chính xác trên những người dùng hay sản phẩm đã có tương tác tương đương với mô hình tiền huấn luyện. Mô hình lấy các tập mẫu ngẫu nhiên từ tương tác quan sát được hay cặp người dùng và sản phẩm có tương tác trong quá khứ.

Để hệ tư vấn vẫn có thể hoạt động tốt khi mà thiếu biểu diễn tương tác của người dùng hay sản phẩm khởi động nguội, với mỗi tập mẫu để huấn luyện mô hình, chọn ra ngẫu nhiên vài cặp rồi điều chỉnh biểu diễn tương tác của người dùng hoặc sản phẩm trong mỗi cặp về không với mục tiêu thúc đẩy mô hình tối ưu hàm O khi không có biểu diễn tương tác:

- Với người dùng khởi động nguội:

$$O_{u,v} = (U_u V_v^T - f(0, \Phi_u^U) f(V_v, \Phi_v^V)^T)^2$$
- Với sản phẩm khởi động nguội:

$$O_{u,v} = (U_u V_v^T - f(U_u, \Phi_u^U) f(0, \Phi_v^V)^T)^2$$

Phương pháp huấn luyện này giúp các điểm dữ liệu ngẫu nhiên có biểu diễn tương tác của người dùng hay sản phẩm bị điều chỉnh về 0 kích lệ mô hình chỉ sử dụng thông tin hỗ trợ để đưa ra đánh giá cũng như tăng sức ảnh hưởng của thông tin hỗ trợ trong đầu ra mô hình và giải quyết vấn đề khởi động nguội. Việc huấn luyện những điểm dữ liệu còn lại khuyến khích mô hình tập trung vào biểu diễn tương tác để đưa ra đánh giá và bỏ qua thông tin hỗ trợ. Tóm lại, mô hình học cách khôi phục độ chính xác của tư vấn từ mô hình được tiền huấn luyện khi có biểu diễn tương tác của người dùng và sản phẩm, đồng thời giải quyết vấn đề khởi động nguội.

Sau khi hoàn thành quá trình huấn luyện là quá trình triển khai mô hình và thu được đầu ra. Trong quá trình triển khai mô hình, những người dùng hay sản phẩm khởi động nguội có khả năng phát sinh tương tác mới, mô hình cần thiết phải cập nhật lại biểu diễn tương tác. Tuy nhiên, huấn luyện mô hình sử dụng những tương tác mới là một bài toán phức tạp, có rủi ro gây ra tình trạng lãng quên tri thức do sự khác nhau về phân phối giữa tương tác mới và tương tác cũ đã được dùng để huấn luyện mô hình trước khi triển khai. Chính vì vậy, tác giả bài báo đã đề xuất sử dụng tổng trung bình biểu diễn tương tác các sản phẩm người dùng khởi động nguội tương tác làm biểu diễn tương tác tạm thời cho họ, tương tự với các sản phẩm khởi động nguội được người dùng tương tác trong quá trình triển khai. Vậy với người dùng khởi động nguội u tương tác với một tập hợp các sản phẩm mới ký hiệu là $V(v)$ trong quá trình triển khai, biểu diễn tương tác tạm thời sẽ được tính theo công thức:

$$U_u = \frac{1}{|V(v)|} \sum_{v \in V(v)} V_v$$

Nhằm giúp phân phối của biểu diễn tương tác tạm thời có sự tương đồng với biểu diễn tương tác của mô hình để ý tưởng nêu bên trên thật sự hiệu quả, trong khi huấn luyện mô hình tối ưu hàm mất mát O , với mỗi tập mẫu dữ liệu, mô hình chọn ra vài điểm dữ liệu và điều chỉnh biểu diễn tương tác người dùng sử dụng làm đầu vào $f(.)$ thành tổng trung bình biểu diễn tương tác các sản phẩm người dùng đó từng tương tác hay ngược lại.

- Với người dùng bị chọn ngẫu nhiên:

$$O_{u,v} = (U_u V_v^T - f(\text{mean}_{v \in V(v)} V_v, \Phi_u^U) f(V_v, \Phi_v^V)^T)^2$$

- Với sản phẩm bị chọn ngẫu nhiên:

$$O_{u,v} = (U_u V_v^T - f(U_u, \Phi_u^U) f(\text{mean}_{u \in U(u)} U_u, \Phi_v^V)^T)^2$$

Toàn bộ quá trình huấn luyện của mô hình có thể mô tả bằng mã giả như sau:

Thuật toán 1: cách thức huấn luyện DropoutNet			
	Đầu vào: R, U, V, Φ^U, Φ^V		
	Đầu ra: f_U, f_V		
	Khởi tạo: Mạng tinh chỉnh học biểu diễn tương tác người dùng là f_U , mạng tinh chỉnh học biểu diễn tương tác sản phẩm là f_V		
	repeat {huấn luyện mô hình}		
		lấy ngẫu nhiên một tập mẫu nhỏ kí hiệu là $B = \{(u_1, v_1), \dots, (u_k, v_k)\}$ với kích thước là k từ R	
		for $(u, v) \in B$ do	
			Thực hiện ngẫu nhiên một trong các lựa chọn sau:
			1. giữ nguyên
			2. điều chỉnh các giá trị trong biểu diễn người dùng về 0. $[U_u, \Phi_u^U] \rightarrow [0, \Phi_u^U]$
			3. điều chỉnh các giá trị trong biểu diễn sản phẩm về 0: $[V_v, \Phi_v^V] \rightarrow [0, \Phi_v^V]$
			4. điều chỉnh biểu diễn của người dùng về tổng trung bình biểu diễn các sản phẩm người dùng đó tương tác: $[U_u, \Phi_u^U] \rightarrow [mean_{v \in V(u)} V_v, \Phi_u^U]$
			5. điều chỉnh biểu diễn của sản phẩm về tổng trung bình biểu diễn các người dùng tương tác sản phẩm đó: $[V_v, \Phi_v^V] \rightarrow [mean_{u \in U(v)} U_u, \Phi_v^V]$
		end for	
		cập nhật f_U, f_V sử dụng B	
	until mô hình hội tụ		

Ưu điểm của mô hình:

- Mô hình xử lý được cả hai vấn đề người dùng và sản phẩm khởi động nguội.
- Duy trì hiệu năng tư vấn trên người dùng cũ.
- Có thể áp dụng cho đa phần các nhóm mô hình tiềm ẩn, giảm thiểu khó khăn về mặt triển khai cho môi trường sản phẩm đã sử dụng các mô hình tiềm ẩn.

Nhược điểm của mô hình:

- Nhiều tác vụ tối ưu khác nhau, chưa có phương thức chất lọc tri thức giảm thiểu rủi ro tới tự sự khác biệt về phân phối của các tác vụ, có thể xảy ra tình trạng quá khớp cũng như thiên vị và giảm hiệu năng tư vấn của mô hình.
- Mô hình bị phụ thuộc vào hiệu năng của mô hình tiền huấn luyện học biểu diễn tương tác của người dùng và sản phẩm.

2.2. Sơ lược về mô hình GIFT4Rec cho khởi đầu nguội

Mô hình GIFT4Rec được đề xuất trong bài “GIFT4Rec: An Effective Side Information Fusion Technique Apply to Graph Neural Network for Cold-Start Recommendation” [17] đã được công bố tại hội nghị ACIIDS vào năm 2023.

Mô hình này có phạm vi áp dụng trong bài toán người dùng khởi động nguội. Mô hình lấy động lực từ khó khăn khi thông tin hỗ trợ được kết hợp không kiểm soát trong hệ tư vấn giải quyết bài toán khởi động nguội được đề cập trong chương 1.2, xảy ra tình trạng quá khớp cũng như giảm hiệu năng tư vấn trên tập người dùng và sản phẩm có tương tác.

Mô hình sử dụng nhóm giải pháp dựa trên mạng nơ-ron đồ thị để học biểu diễn tương tác cũng như ánh xạ thông tin hỗ trợ của người dùng qua mạng nơ-ron nhân tạo để đưa về véc tơ cùng độ dài và phân phối với biểu diễn tương tác. Để giải quyết tình trạng thông tin hỗ trợ được kết hợp vào hệ tư vấn không kiểm soát cũng như duy trì hiệu năng tư vấn trên người dùng đã có tương tác để hướng đến hệ tư vấn công bằng, mô hình đề xuất 1 phương thức học trọng số cho biểu diễn tương tác và thông tin hỗ trợ của người dùng. Trọng số này được học qua hai hàm mất mát phục vụ hai tác vụ khác nhau nâng cao hiệu năng tư vấn của bài toán.

Ưu điểm:

- Mô hình giải quyết được bài toán người dùng khởi động nguội nhưng vẫn duy trì được hiệu năng tư vấn trên người dùng cũ.
- Mô hình học được trọng số để kết hợp biểu diễn tương tác và thông tin hỗ trợ người dùng hợp lý.
- Mô hình đề xuất giải pháp huấn luyện trên đa tác vụ tối ưu hiệu năng tư vấn trên các góc nhìn khác nhau.

Nhược điểm:

- Hiệu năng cải tiến chưa đáng kể.

- Nhiều tác vụ tối ưu khác nhau, chưa có phương thức chắt lọc tri thức giảm thiểu rủi ro tới tự sự khác biệt về phân phối của các tác vụ, gây ra tình trạng quá khớp cũng như thiên vị và giảm hiệu năng tư vấn của mô hình. Vấn đề này cũng được nêu ra trong phần đánh giá kết quả của mô hình trong bài báo.

2.3. Ý tưởng cải tiến của khóa luận

Sau quá trình phân tích và nghiên cứu GIFT4Rec [17], khóa luận nhận thấy các giới hạn sau đây:

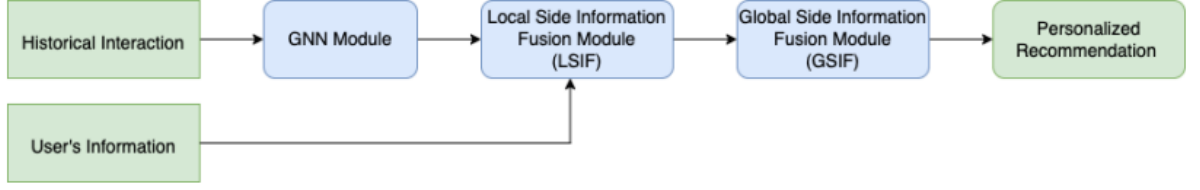
- Tuy đã có cơ chế học trọng số kết hợp biểu diễn tương tác và thông tin hỗ trợ của người dùng, nhưng mô hình học đồng thời mạng ánh xạ biểu diễn từ tương tác và mạng trích xuất hành vi từ thông tin hỗ trợ trong cùng một lần lan truyền ngược. Điều này cũng gây ra tình trạng quá khớp cho mạng ánh xạ biểu diễn từ tương tác bị ảnh hưởng từ nhiễu của thông tin hỗ trợ khi chúng không có ý nghĩa trong việc đánh giá hành vi người dùng. Để khắc phục vấn đề này, khóa luận đề xuất một cơ chế huấn luyện có tham khảo ý tưởng một phần từ mô hình DropoutNet [10], đó là tiền huấn luyện mạng ánh xạ biểu diễn từ tương tác trước và đóng băng chúng khi học để kết hợp thêm thông tin hỗ trợ. Mục đích để giúp quá trình học biểu diễn qua tương tác trọn vẹn và độc lập nhất, nhưng vẫn bổ sung tri thức cho quá trình trích xuất hành vi từ thông tin hỗ trợ khi học kết hợp, nâng cao khả năng nắm bắt sở thích người dùng của mô hình
- Bài toán học đa tác vụ với phân phối khác nhau và vấn đề gây ảnh hưởng không tốt đến kết quả, điều này đã được đề cập đến trong phần nhận xét kết quả thực nghiệm trong bài báo. Khóa luận sẽ đề xuất một giải pháp bắt nguồn trong lớp bài toán học liên tục (Continual Learning) để chắt lọc được tri thức cần và bảo vệ các tri thức quan trọng của mỗi tác vụ, giúp mô hình học đồng thời các tác vụ hiệu quả hơn.

Kết luận

Trong chương 2, khóa luận đã trình bày về các giải pháp giải quyết vấn đề khởi động nguội nhưng vẫn đảm bảo hiệu năng tư vấn trên người dùng cũ. Lấy những nhược điểm còn tồn đọng trong các lớp giải pháp, khóa luận đã trình bày về ý tưởng cải tiến. Trong chương 3, khóa luận sẽ trình bày cụ thể về mô hình đề xuất

CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT

3.1. Mô hình GIFT4Rec



Hình 3.1: Kiến trúc tổng quan của mô hình GIFT4Rec

Kiến trúc tổng quan của mô hình được thể hiện trong hình 3.1 trích dẫn từ bài nghiên cứu gốc [17]. Mô hình bao gồm hai thành phần: mô-đun mạng nơ-ron đồ thị (GNN) và mô-đun kết hợp thông tin hỗ trợ cục bộ và toàn cục. Mô-đun GNN tìm hiểu và trích xuất các đặc điểm về hành vi của người dùng và cách trình bày của mục. Dữ liệu tương tác giữa người dùng và sản phẩm được biểu diễn dưới dạng đồ thị hai phía. Biểu diễn tương tác của người dùng và sản phẩm được ánh xạ qua các lớp của mô-đun mạng nơ-ron đồ thị. Sau đó, mô-đun kết hợp thông tin hỗ trợ học cách kết hợp thông tin hỗ trợ với biểu diễn tương tác của người dùng là đầu ra của mô-đun mạng nơ-ron đồ thị.

Biểu diễn tương tác của người dùng i và sản phẩm j từ mô-đun này được kí hiệu là X_{u_i} và X_{i_j} . Biểu diễn cho thông tin hỗ trợ đã được ánh xạ về cùng phân phối và độ dài với biểu diễn tương tác của người dùng i được kí hiệu là: $X_{info_{u_i}}$. Biểu diễn sau cùng cho người dùng i sau khi kết hợp X_{u_i} và X_{i_j} được kí hiệu là $X_{final_{u_i}}$. Công thức tính khả năng người dùng i tương tác sản phẩm j mô hình sử dụng là:

$$y_{u_i, i_j} = X_{final_{u_i}} X_{i_j}^T$$

Mô hình sẽ tối ưu y_{u_i, i_j} sử dụng tương tác của người dùng và sản phẩm đã quan sát được qua hàm mất mát BPR:

$$L_{CF} = - \sum_{(u_i, i_j, i_j')} \log(y_{u_i, i_j} - y_{u_i, i_j'})$$

Với u_i, i_j lần lượt là hai cặp người dùng và sản phẩm đã có tương tác với nhau, i_j' được ngẫu nhiên lựa chọn từ danh sách sản phẩm u_i chưa tương tác làm mẫu âm.

Mô-đun mạng nơ-ron nhân tạo:

Danh sách sản phẩm trong bài toán là $V = \{v_1, v_2, \dots, v_p\}$ cùng số lượng là p . Với người dùng $u_i, i \in \{1, 2, \dots, N\}$ với thông tin hỗ trợ là X_{info_i} , danh sách sản phẩm người dùng này tương tác là $S_i = \{s_{i1}, s_{i2}, \dots, s_{iq}; s_{ij} \in T, q \leq p\}$

Đồ thị sử dụng được biểu diễn dưới dạng $G = (U, V)$ hay $\{(u_i, s_{ij}, v_j) \mid u_i \in U, v_j \in V\}$ với U, V lần lượt là tập người dùng, tập sản phẩm đã được định nghĩa bên trên và liên kết $s_{ij} = 1$ đại diện cho tương tác của người dùng i và sản phẩm j . lân cận của một nút trong đồ thị là thực thể người dùng hay sản phẩm được ký hiệu là $N(\cdot)$. Mô-đun mạng nơ-ron đồ thị xếp chồng nhiều lớp lan truyền, bao gồm các công việc chính là tổng hợp và cập nhật:

$$\text{Tổng hợp: } n^{(l)} = \text{Tổng hợp}_l(\{h_u^l, \forall u \in N(\cdot)\})$$

$$\text{Cập nhật: } h^{(l+1)} = \text{Cập nhật}_l(h^{(l)}, n^{(l)})$$

Với $h_{u_i}^l$ và $h_{v_j}^l$ đại diện cho biểu diễn tương tác của người dùng i và sản phẩm j tại tầng thứ l của mô-đun mạng nơ-ron đồ thị. Tổng hợp_l và Cập nhật_l đại diện cho các hàm tổng hợp và cập nhật tại tầng thứ l . Giả như mô-đun này có n tầng thì biểu diễn tương tác của người dùng i và sản phẩm j sẽ được tính theo công thức:

$$X_{u_i} = h_{u_i}^n$$

$$X_{i_j} = h_{i_j}^l$$

Mô-đun kết hợp thông tin bổ trợ:

Mô-đun sử dụng hai kỹ thuật kết hợp thông tin bổ trợ, quan sát người dùng từ các góc độ khác nhau để nâng cao hiệu năng tư vấn. Mô-đun này nhằm mục đích kiểm soát và ước lượng trọng số của từng nguồn thông tin (tương tác với sản phẩm và thông tin bổ trợ) để kết hợp chúng hiệu quả và mang lại sự công bằng khi tư vấn. Thành phần đầu tiên là mô-đun kết hợp thông tin bổ trợ cục bộ huấn luyện mô-đun sinh trọng số qua hàm mất mát L_{CF} . Thành phần còn lại giảm tình trạng quá khớp vào dữ liệu huấn luyện trong quá khứ và giúp mô hình nắm bắt những tri thức mang tính tổng quan hơn. Hai mô-đun này chia sẻ các tham số từ mạng tạo ra trọng số cho thông tin bổ trợ và biểu diễn tương tác của người dùng được gọi là mô-đun sinh ra trọng số.

Trước tiên là mô-đun kết hợp thông tin bổ trợ cục bộ. Trong mô-đun này, bài báo đã đề xuất một thuật toán mới mang tên là ADN sử dụng kỹ thuật trong mô hình DropoutNet [10] trong mô-đun sinh trọng số để kiểm soát thông tin bổ trợ và biểu diễn tương tác mỗi người dùng tốt hơn. Cách thức mô-đun sinh trọng số hoạt động như hình mô tả:

- Nối biểu diễn tương tác và biểu diễn thông tin bổ trợ của người dùng i :

$$X_{concat_{u_i}} = \text{concat}([X_{u_i}, X_{info_{u_i}}])$$

- Mô-đun sinh trọng số sử dụng mạng nơ-ron nhân tạo như 1 hàm ánh xạ đặt tên là $WG(\cdot)$ để ánh xạ $X_{concat_{u_i}}$ tới trọng số a_{u_i} của X_{u_i} và $1 - a_{u_i}$ là trọng số của $X_{info_{u_i}}$:

$$a_{u_i} = WG(X_{concat_{u_i}})$$

- Biểu diễn cuối cùng của người dùng i sẽ được tính toán theo công thức:

$$X_{final_{u_i}} = a_{u_i} \circ X_{u_i} + (1 - a_{u_i}) \circ X_{info_{u_i}}$$

Mô-đun sinh ra trọng số này sẽ được huấn luyện đồng thời cùng các thành phần còn lại khi tối ưu hàm mất mát L_{CF} . Để tăng cường vai trò của thông tin hỗ trợ trong đầu ra mô hình, giải quyết vấn đề khởi động nguội. Trong khi huấn luyện, với một tập mẫu dữ liệu, chọn ngẫu nhiên vài điểm dữ liệu loại bỏ biểu diễn tương tác và chỉ đưa thông tin hỗ trợ vào mô hình:

$$X_{final_{u_i}} = X_{info_{u_i}}$$

Trong quá trình triển khai mô hình, biểu diễn tương tác của những người dùng khởi động nguội sẽ tạm thời được tính toán bằng tổng trung bình biểu diễn tương tác của những người dùng đã có tương tác.

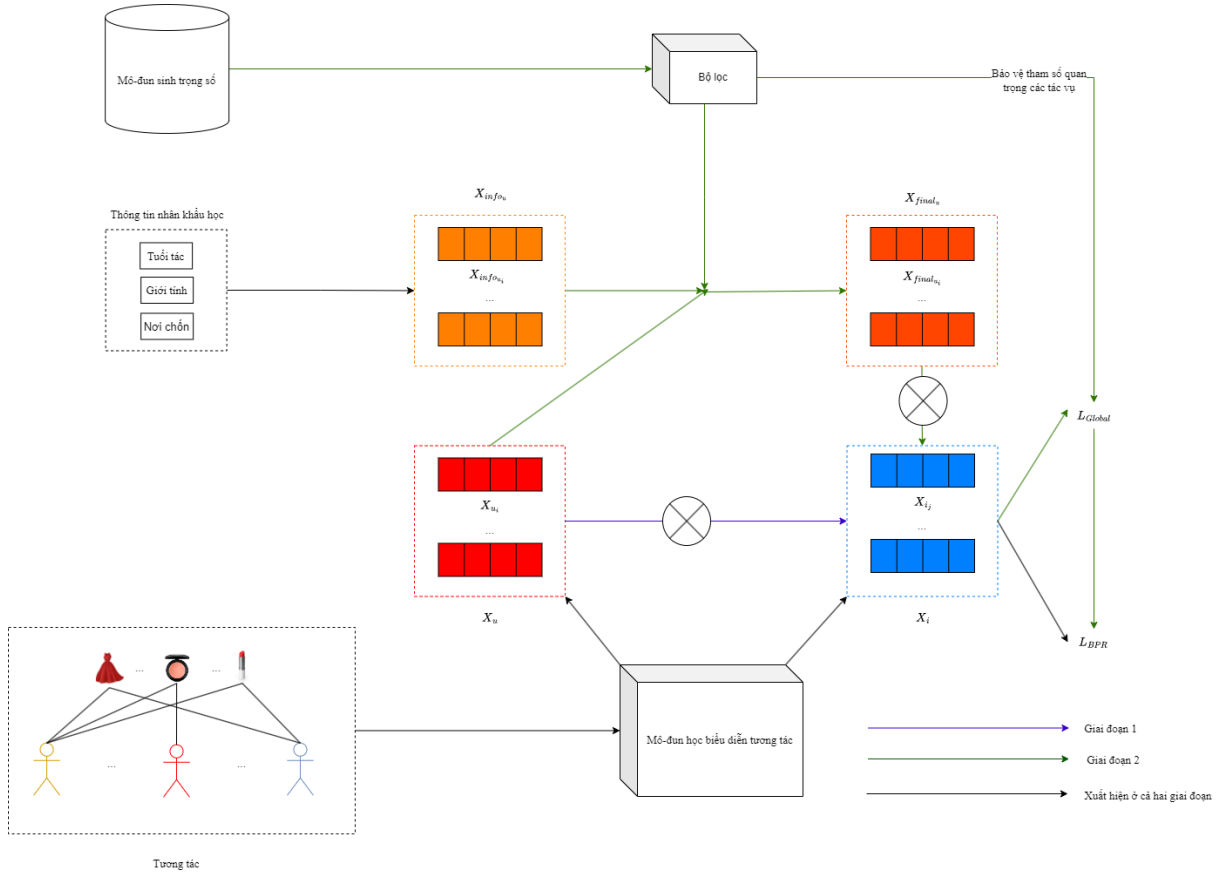
Tiếp theo là mô-đun kết hợp thông tin hỗ trợ toàn cục. Để giảm tình trạng quá khớp của mô hình và giúp mô hình thích nghi nhanh hơn trên tác vụ và dữ liệu mới, bài báo đề xuất một kỹ thuật siêu học dựa trên độ đo thông qua so sánh hiệu năng của mô hình nếu chỉ dùng biểu diễn tương tác hoặc thông tin hỗ trợ trên tập dữ liệu con không quan sát được khi huấn luyện. Với mỗi người dùng i , bài báo định nghĩa $label_{u_i} = 1$ nếu hiệu năng khi chỉ sử dụng biểu diễn tương tác tốt hơn, ngược lại thì $label_{u_i} = 0$. Để mô-đun sinh trọng số có thể tối ưu trực tiếp những độ đo hiệu năng của hệ tư vấn mà L_{CF} không làm được, bài báo đề xuất tối ưu trọng số đầu ra của mô-đun này thông qua nhãn $label_{u_i}$ bằng 1 hàm mất mát Cross-Entropy được định nghĩa là L_{Global} :

$$L_{Global} = - \sum_{i \in U} label_{u_i} \log(a_{u_i}) + (1 - label_{u_i}) \log(1 - a_{u_i})$$

Trong quá trình tối ưu hàm mất mát L_{Global} , tham số của tất cả thành phần trừ mô-đun sinh trọng số sẽ bị đóng băng để giảm thời gian huấn luyện. Giải pháp này giảm thiểu tình trạng mô hình quá khớp với các tương tác trong quá khứ và trực tiếp tối ưu hiệu suất mô hình được tính toán bằng các độ đo kinh doanh mà không thể được tối ưu trực tiếp bằng các hàm mất mát thông thường. Mô hình sẽ được học một cách hiệu quả hơn việc chỉ tập trung huấn luyện chúng dự đoán tốt hơn các tương tác đã quan sát được trong quá khứ.

3.2. Mô hình cải tiến

Khóa luận đề xuất mô hình với kiến trúc tổng quan như hình 3.2 dưới đây:



Hình 3.2: Kiến trúc tổng quan của mô hình cải tiến

Nhược điểm của GIFT4Rec [17] đã được mô tả sơ lược trong chương 2. Trong chương này khóa luận sẽ trình bày cụ thể hơn về các nhược điểm còn tồn đọng trong mô hình, từ đây sử dụng làm động lực và đề xuất giải pháp khắc phục cũng như cải tiến.

Trước hết, khóa luận trình bày chi tiết về nhược điểm đầu tiên. Tuy đã có mô-đun sinh trọng số để kiểm soát lượng sử dụng từ biểu diễn tương tác với thông tin hỗ trợ để kết hợp tăng tính công bằng trong hệ tư vấn, nhưng mô-đun này sẽ chỉ hoạt động tốt nếu như hai loại biểu diễn cũng được học chính xác. GIFT4Rec huấn luyện đồng thời hai thành phần là mô-đun mạng nơ-ron đồ thị và mạng trích xuất thông tin hỗ trợ trong quá trình lan truyền ngược. Với kiến trúc mô hình và cách thức huấn luyện này, mô-đun học biểu diễn tương tác bị ảnh hưởng xấu bởi nhiễu đến từ thông tin hỗ trợ khi chúng không thật sự có ích để đánh giá sở thích của người dùng, từ đây gặp phải tình trạng quá khớp cũng như không thể tìm ra biểu diễn đúng của người dùng, ảnh hưởng trực tiếp đến hiệu năng tư vấn trên người dùng cũ.

Để khắc phục nhược điểm này, quá trình huấn luyện mô hình chia làm hai giai đoạn được mô tả ở hình vẽ 3.2:

- Giai đoạn đầu tiên sẽ chỉ tập trung huấn luyện trước mô-đun học biểu diễn tương tác qua hàm mất mát L_{CF} một cách độc lập mà không sử dụng thêm thông tin hỗ trợ. Với mỗi người dùng i , mô hình sử dụng X_{u_i} làm biểu diễn tương tác cuối cùng của họ:

$$X_{final_{u_i}} = X_{u_i}$$

- Giai đoạn thứ hai sẽ huấn luyện mô hình qua hai hàm mất mát là L_{CF} và L_{Global} , kết hợp thông tin hỗ trợ với biểu diễn tương tác của người dùng. Trong giai đoạn thứ hai, mô-đun học biểu diễn tương tác hoàn toàn sẽ được đóng băng để giảm thiểu từ thông tin hỗ trợ.

Cách thức huấn luyện này giúp mô-đun mạng nơ-ron đồ thị có thể học biểu diễn từ tương tác một cách trọn vẹn nhất và loại bỏ nhiễu có thể xuất phát từ thông tin hỗ trợ, nhưng vẫn sử dụng tri thức từ tương tác người dùng để có thể học biểu diễn thông tin hỗ trợ cũng như trọng số kết hợp tốt hơn. Cách thức huấn luyện này được tham khảo 1 phần từ mô hình DropoutNet [10], tuy nhiên cách thức huấn luyện của khóa luận đề xuất khắc phục tình trạng bị phụ thuộc vào hiệu năng của mô-đun học biểu diễn tương tác, bảo toàn biểu diễn tương tác đã được tiền huấn luyện và tối ưu trực tiếp vào các độ đo hiệu năng bài toán.

Nhược điểm thứ hai của GIFT4Rec nằm ở việc huấn luyện đồng thời đa tác vụ. Có hai nhiệm vụ khác nhau là tối ưu hàm mất mát L_{CF} và L_{Global} , tuy nhiên chưa có phương thức chất lọc tri thức giảm thiểu rủi ro tới sự khác biệt về phân phối của các tác vụ, gây ra tình trạng quá khớp cũng như thiên vị và giảm hiệu năng tư vấn của mô hình. Vấn đề này cũng được nêu ra trong phần đánh giá kết quả của mô hình trong bài báo.

Lấy nhược điểm trên làm động lực, trong giai đoạn thứ hai huấn luyện mô hình để kết hợp thông tin hỗ trợ với biểu diễn tương tác, với mục tiêu giảm thiểu ảnh hưởng xấu từ sự khác biệt về phân phối của L_{CF} và L_{Global} , khóa luận đề xuất một kỹ thuật bắt nguồn từ các lớp mô hình trong nhóm bài toán học liên tục để chất lọc tri thức quan trọng trong từng tác vụ và bảo vệ chúng khỏi tác động xấu khi học các tác vụ khác. Với mỗi tác vụ, khóa luận sẽ đưa ra 1 bộ lọc để có thể tập trung sử dụng nhiều hơn những tham số có lợi khi đánh giá và bảo vệ chúng trong quá trình học các tác vụ khác. Khóa luận kí hiệu tham số của tầng thứ n trong mô-đun sinh ra trọng số của mô hình dựa trên mạng nơ-ron nhân tạo lần lượt là W_n với b_n . Hai bộ lọc cho tầng thứ n với nhiệm vụ k sẽ lần lượt được kí hiệu $mask_{n_{W_k}}$ với $mask_{n_{b_k}}$.

Bộ trọng số mới cho tầng thứ n của nhiệm vụ được định danh là k sẽ được tính theo công thức sau:

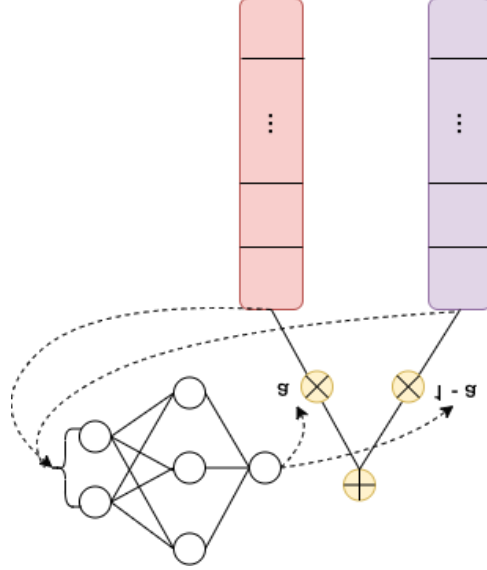
$$\begin{aligned} W_n^k &= W_n \circ \text{sigmoid}(mask_{n_{W_k}}) \\ b_n^k &= b_n \circ \text{sigmoid}(mask_{n_{b_k}}) \end{aligned}$$

Với $\text{sigmoid}(\cdot)$ là hàm sigmoid.

Kí hiệu $f_n(.)$ là hàm kích hoạt của tầng thứ n của mô-đun sinh trọng số. Đầu ra tầng này sẽ được tính theo công thức như sau:

$$\alpha_{u_{in}} = f_m(\alpha_{u_{in-1}} W_n^k + b_n^k)$$

Mô-đun sinh trọng số có thể được mô tả trực quan như hình 3.3 trích dẫn từ bài báo đề xuất mô hình GIFT4Rec [17]:



Hình 3.3: Mô-đun sinh trọng số

Kỹ thuật sinh ra bộ tham số mới này giúp mô hình tập trung vào những tham số quan trọng của tác vụ hơn, giảm thiểu có thể gây ra từ các tham số kém quan trọng dẫn đến tình trạng quá khớp.

Kí hiệu K là tập hợp các tác vụ phải học. Trong quá trình học nhiệm vụ $k \in K$, đạo hàm của bộ trọng số W_n và b_n sẽ được tính toán lại dưới công thức sau đây, trước khi được sử dụng cập nhật lại trọng số, để cân bằng giữa việc giảm ảnh hưởng xấu tới những trọng số quan trọng của nhiệm vụ khác và chuyển giao tri thức để học nhiệm vụ k tốt hơn. Khóa luận kí hiệu đạo hàm của của nhiệm vụ k cho W_n với b_n lần lượt là $G_{W_n^k}$ và $G_{b_n^k}$. Vậy công thức tính đạo hàm mới là:

$$G'_{W_n^k} = G_{W_n^k} \circ \left[1 - \max_pooling \left(\left\{ sigmoid \left(mask_{n_{W_{k_{other}}}} \right) \mid k_{other} \neq k \right\} \right) \right]$$

$$G'_{b_n^k} = G_{b_n^k} \circ \left[1 - \max_pooling \left(\left\{ sigmoid \left(mask_{n_{b_{k_{other}}}} \right) \mid k_{other} \neq k \right\} \right) \right]$$

Với $\max_pooling(.)$ là hàm lấy ra giá trị lớn nhất ở các vị trí tương đương nhau tại tất cả các ma trận trong tập hợp chỉ định và tạo thành 1 ma trận với không gian tương đương.

Với công thức bên trên, độ lớn đạo hàm mới của mỗi trọng số sẽ được quyết định dựa vào nhiệm vụ khác mà chúng có vai trò quan trọng nhất. Đạo hàm mới của những trọng số giá trị lớn hơn tại vị trí tương ứng trong tập hợp bộ lọc tổng hợp từ tất cả các tác vụ còn lại sẽ bị giảm đi nhiều hơn, điều này giúp những trọng số quan trọng nhất của từng nhiệm vụ ít bị thay đổi hơn khi mô hình học nhiệm vụ khác nhưng vẫn đảm bảo sự chuyển giao tri thức giữa các nhiệm vụ để nâng cao hiệu năng.

Sau khi quá trình huấn luyện kết thúc và đến quá trình triển khai mô hình, các bộ tham số của mô-đun sinh trọng số tại tầng thứ n sẽ được tính toán bằng công thức như dưới đây mô tả:

$$W'_n = W_n \circ \max_pooling \left(\left\{ sigmoid \left(mask_{n_{w_k}} \right) \mid k \in K \right\} \right)$$

$$b'_n = b_n \circ \max_pooling \left(\left\{ sigmoid \left(mask_{n_{b_k}} \right) \mid k \in K \right\} \right)$$

Dưới đây là mã giả cho quá trình huấn luyện của mô hình cải tiến của khóa luận. Kí hiệu θ , $\theta_{interaction}$, θ_{info} và θ_{WG} lần lượt là tham số học của toàn bộ mô hình, mô-đun học biểu diễn tương tác, mô-đun học biểu diễn hành vi và mô-đun sinh trọng số. R_{train} , R_{val} lần lượt được kí hiệu là tập tương tác sử dụng để huấn luyện mô hình thông qua tối ưu L_{CF} và L_{global} . R_{train} chứa các cặp người dùng i và sản phẩm j có tương tác là (u_i, i_j) . Mỗi phần tử trong R_{val} có dạng là (u_i, i_{u_i}) là cặp người dùng i cùng danh sách sản phẩm người đó tương tác. $limit$ là giới hạn cho việc chỉ dùng thông tin hỗ trợ được đề cập trong chương 3.1.

Thuật toán 2: Huấn luyện mô hình cải tiến của khóa luận		
	Đầu vào: $R_{val}, R_{train}, limit$	
	Đầu ra: θ	
	Khởi tạo: θ	
	$count \leftarrow 0$	
	repeat {huấn luyện mô hình}	
		Lấy tập mẫu B từ R_{train}
		for $(u_i, i_j) \in B$
		X_{u_i} được lấy từ mô-đun học biểu diễn tương tác
		$X_{final_{u_i}} \leftarrow X_{u_i}$
		Lấy ngẫu nhiên một mẫu sản phẩm i_j , mà u_i chưa từng tương tác.
		$y_{u_i.i_j} \leftarrow X_{final_{u_i}} X_{i_j}^T$
		$y_{u_i.i_{j'}} \leftarrow X_{final_{u_i}} X_{i_{j'}}^T$
		end for
		Tối ưu L_{CF} và cập nhật $\theta_{interaction}$
		$k \leftarrow 0$
		Lấy tập mẫu B từ R_{train}
		for $(u_i, i_j) \in B$
		X_{u_i} được lấy từ mô-đun học biểu diễn tương tác
		$X_{concat_{u_i}} \leftarrow concat([X_{u_i}, X_{info_{u_i}}])$
		$\alpha_{u_{i_0}} \leftarrow X_{concat_{u_i}}$
		for $m = 1$ to n
		$\alpha_{u_{i_m}} \leftarrow f_m(\alpha_{u_{i_{m-1}}} W_m^k + b_m^k)$
		end for
		$\alpha_{u_i} \leftarrow \alpha_{u_{i_n}}$
		rate được lấy ngẫu nhiên từ $[0, 1]$
		$X_{final_{u_i}} \leftarrow \alpha_{u_i} \circ X_{u_i} + (1 - \alpha_{u_i}) \circ X_{info_{u_i}}$
		if rate $\leq limit$
		then $X_{final_{u_i}} \leftarrow X_{info_{u_i}}$
		Lấy ngẫu nhiên một mẫu sản phẩm i_j , mà u_i chưa từng tương tác.
		$y_{u_i.i_j} \leftarrow X_{final_{u_i}} X_{i_j}^T$
		$y_{u_i.i_{j'}} \leftarrow X_{final_{u_i}} X_{i_{j'}}^T$
		end for
		Tính toán đạo hàm của L_{CF}
		length = n
		if count = 0

			then length = 0
			for $m = 1$ to length
			$G'_{b_m^0} \leftarrow G_{b_m^0} \circ \left(1 - \text{sigmoid}(\text{mask}_{m_{b_1}})\right)$
			$G'_{W_m^0} \leftarrow G_{W_m^0} \circ \left(1 - \text{sigmoid}(\text{mask}_{m_{W_1}})\right)$
			end for
			if length = n
			then Cập nhật θ_{info} và θ_{WG} sử dụng đạo hàm mới
			else Cập nhật θ_{info} và θ_{WG} sử dụng đạo hàm cũ
			$k \leftarrow 1$
			Lấy tập mẫu B từ R_{val}
			for $(u_i, i_{u_i}) \in B$
			X_{u_i} được lấy từ mô-đun học biểu diễn tương tác
			$X_{concat_{u_i}} \leftarrow \text{concat}([X_{u_i}, X_{info_{u_i}}])$
			$\alpha_{u_{i_0}} \leftarrow X_{concat_{u_i}}$
			for $m = 1$ to n
			$\alpha_{u_{i_m}} \leftarrow f_m(\alpha_{u_{i_{m-1}}} W_m^k + b_m^k)$
			end for
			$\alpha_{u_i} \leftarrow \alpha_{u_{i_n}}$
			for $i_j \in I$
			$y_{u_i.i_j} \leftarrow X_{u_i} X_{i_j}^T$
			$y_{info_{u_i}.i_j} \leftarrow X_{info_{u_i}} X_{i_j}^T$
			Gán $label_{u_i}$ dựa vào i_{u_i} và $\{y_{u_i.i_j}, y_{info_{u_i}.i_j} i_j \in I\}$
			end for
			Tính đạo hàm của L_{global}
			for $m = 1$ to n
			$G'_{b_m^1} \leftarrow G_{b_m^1} \circ \left(1 - \text{sigmoid}(\text{mask}_{m_{b_0}})\right)$
			$G'_{W_m^1} \leftarrow G_{W_m^1} \circ \left(1 - \text{sigmoid}(\text{mask}_{m_{W_0}})\right)$
			end for
			Cập nhật θ_{WG} sử dụng đạo hàm mới
			until mô hình hội tụ

Kết luận

Trong chương 3, khóa luận đã trình bày được cụ thể về mô hình đề xuất, chỉ ra các nhược điểm còn tồn tại của mô hình GIFT4Rec [17] và giải pháp khắc phục. Trong chương tiếp theo, khóa luận sẽ trình bày về quá trình thực nghiệm chứng minh hiệu năng và tính đúng đắn của mô hình đề xuất.

CHƯƠNG 4. THỰC NGHIỆM VÀ ĐÁNH GIÁ

4.1. Phần cứng và phần mềm

Phần cứng

Quá trình thực nghiệm được thực hiện hoàn toàn trên máy tính cá nhân và trang web Google Colab:

	Máy tính cá nhân	Google Colab
Nguồn		https://colab.research.google.com/
Loại GPU	Nvidia RTX 3060	Nvidia K80 / T4
Dung lượng GPU	8GB	12GB / 16GB
Dung lượng Ram	16GB	12GB
Dung lượng bộ nhớ	134GB	358 GB

Bảng 4.1: *Phần cứng sử dụng trong quá trình thực nghiệm*

Phần mềm

Khóa luận sử dụng và phát triển mã nguồn khóa luận từ mã nguồn trong link github: <https://github.com/LunaBlack/KGAT-pytorch>. Đây là mã nguồn mở chứa nhiều thuật toán hệ tư vấn dựa trên đồ thị và được đánh giá nhiều sao trên github.

Sau đây là link mã nguồn của khóa luận được upload trên trang web google drive: https://github.com/darkreaper123/do_an

4.2. Các kịch bản thực nghiệm

Để đánh giá hiệu năng của mô hình cải tiến, khóa luận thực hiện hai thí nghiệm sau:

- So sánh và đánh giá hiệu năng mô hình với các giải pháp hiện có.
- Chứng minh sự quan trọng của từng thành phần trong mô hình qua thí nghiệm cắt bỏ thành phần của mô hình cải tiến và đánh giá kết quả.

4.3. Dữ liệu thử nghiệm

Nhóm nghiên cứu thực nghiệm trên bộ dữ liệu Douban và Movielens 1M. Đây là hai bộ dữ liệu nổi tiếng thường được sử dụng nghiên cứu hệ tư vấn. Nhóm nghiên cứu tiến hành chia dữ liệu như sau:

- Thí nghiệm loại bỏ những sản phẩm người dùng tương tác mà có điểm số được đánh giá nhỏ hơn 3.
- 10% người dùng có ít tương tác nhất sẽ được sử dụng làm tập người dùng khởi động nguội, còn lại là người dùng đã có tương tác.

- Tiến hành loại bỏ hết tương tác của những người dùng mới trong tập huấn luyện và sử dụng làm tập kiểm tra cho người dùng khởi động nguội. Với những người dùng cũ thì tương tác cuối cùng của họ sẽ loại bỏ trong tập huấn luyện và sử dụng làm tập kiểm tra cho người dùng đã có tương tác.
- Tiếp tục tiến hành loại bỏ những tương tác cuối cùng trong tập huấn luyện để làm tập kiểm định. Sử dụng cho mô hình GIFT4Rec [17] và mô hình đề xuất huấn luyện tác vụ thứ hai.

4.4. Độ đo

Khóa luận sử dụng một độ đo được sử dụng để tính hiệu năng tư vấn được định nghĩa là $Recall@k$ với k là số nguyên dương định trước. Độ đo này có công thức tính như sau:

$$Recall@k = \sum_{u \in U} \frac{|S\grave{a}n\ ph\grave{a}m\ g\grave{o}i\ y\ \acute{d}\acute{u}ng\ t\grave{u}r\ danh\ s\acute{a}ch\ k\ s\grave{a}n\ ph\grave{a}m\ tư\ v\grave{a}n|}{|S\grave{a}n\ ph\grave{a}m\ th\acute{u}c\ t\acute{e}\ ng\ddot{u}oi\ d\grave{u}ng\ t\grave{u}ng\ t\acute{a}c|}$$

Khóa luận sẽ sử dụng độ đo này với $k \in \{20, 50\}$ trên từng tập dữ liệu kiểm tra của người dùng đã có tương tác và khởi động nguội.

Để đảm bảo tính công bằng của hệ tư vấn giữa người dùng đã có tương tác và khởi động nguội. Khóa luận đề xuất sử dụng thêm 2 độ đo như sau. Định nghĩa tập người dùng trong tập dữ liệu kiểm tra của người dùng đã có tương tác và khởi động nguội lần lượt là U_{test} và $U_{test\ cold-start}$ cùng độ đo hiệu năng tư vấn mỗi người dùng là $R(.)$:

- Độ đo thứ nhất khóa luận sử dụng là GF được đề xuất tại [24], đại lượng này đã được đề cập trong chương 1.3. Đại lượng này tính toán độ chênh lệch giữa hiệu năng tư vấn các nhóm người dùng trong bài toán định nghĩa để đánh giá sự công bằng. Công thức sử dụng trong bài toán của khóa luận sẽ được định nghĩa như sau:

$$GF = \left| \frac{1}{|U_{test}|} \sum_{u_1 \in U_{test}} R(u_1) - \frac{1}{|U_{test\ cold-start}|} \sum_{u_2 \in U_{test\ cold-start}} R(u_2) \right|$$

- Tuy độ đo thứ nhất đã thể hiện được tính công bằng của hệ tư vấn, nhưng lại không thể hiện được độ lớn của hiệu năng tư vấn trên các nhóm người dùng. Vậy nên khóa luận đề xuất độ đo mới là tổng trung bình hiệu năng tư vấn trên các nhóm người dùng, cân bằng giữa độ lớn và độ chênh lệch hay tính công bằng của hiệu năng tư vấn các nhóm:

$$GF_2 = \frac{\left(\frac{1}{|U_{test}|} \sum_{u_1 \in U_{test}} R(u_1) + \frac{1}{|U_{test\ cold-start}|} \sum_{u_2 \in U_{test\ cold-start}} R(u_2) \right)}{2}$$

- Khóa luận kí hiệu $GF_{2R(.)}$ Và $GF_{R(.)}$ lần lượt đại diện cho GF_2 và GF khi sử dụng $R(.)$ làm phương thức đánh giá hiệu năng. Như đã nêu bên trên, trong thí nghiệm của khóa luận $R(.)$ sẽ là $Recall@20$ hoặc $Recall@50$.

4.5. Kết quả thực nghiệm

Thí nghiệm thứ nhất để so sánh hiệu năng mô hình cải tiến với giải pháp đã có. Thí nghiệm thứ nhất sử dụng mô hình DropoutNet [10] đã được trình bày ở chương 2 của khóa luận để so sánh cũng như mô hình GIFT4Rec [17] để đánh giá hiệu năng của mô hình trước và sau khi cải tiến của khóa luận. Bên cạnh đó, khóa luận sử dụng thêm mô hình MeLU [25] để so sánh với phương pháp siêu học giải quyết vấn đề khởi động nguội. Bài báo đề xuất mô hình không liên hệ trực tiếp đến đề tài khóa luận, nhưng trong phần thực nghiệm có đề cập và so sánh hiệu năng tư vấn của người dùng đã có tương tác và cho thấy kết quả tốt. Tuy nhiên, cách thức đánh giá hiệu năng mô hình trong bài báo thiếu tính thực tế, chỉ đánh giá thứ tự xếp hạng của mô hình trong tập sản phẩm tương lai người dùng tương tác. Trên thực tế, các mô hình tư vấn sẽ xếp hạng lại toàn bộ danh sách sản phẩm người dùng chưa tương tác và chọn ra các sản phẩm mới phù hợp nhất, đây cũng chính là cách thức đánh giá hiệu năng tư vấn của khóa luận sử dụng trong hai thí nghiệm trình bày. Vậy nên, trong thí nghiệm thứ nhất, khóa luận đã thực nghiệm lại mô hình MeLU sử dụng mã nguồn và siêu tham số của bài báo gốc cũng như chung tập dữ liệu Movielens 1M.

Dữ liệu	Độ đo		MeL U	Dropout- Net	GIFT4Re c	Mô hình cải tiến
Movielens 1M	U_{test}	Recall@20	0.626	6,413	6,35	7,857
		Recall@50	1.748	14,083	12,73	15,927
	$U_{testcold-start}$	Recall@20	0.714	10,916	13,71	13,806
		Recall@50	1.735	23,237	25,876	24,997
		$GF_{2Recall@20}$	0.67	8,67	10,03	10,83
		$GF_{2Recall@50}$	1.741	18,66	19,306	20,463
		$GF_{Recall@20}$	0.088	4,503	7,36	5,949
		$GF_{Recall@50}$	0.013	9,154	13,146	9,07
Douban	U_{test}	Recall@20		2,91	3.56	9,17
		Recall@50		10,93	12.98	16,18
	$U_{testcold-start}$	Recall@20		2,78	5.63	7,41
		Recall@50		11,53	13.29	15,94
		$GF_{2Recall@20}$		2,845	4.6	8,29

		$GF_2_{\text{Recall@50}}$		11,23	13.13	16,06
		$GF_{\text{Recall@20}}$		0,13	2.07	1,76
		$GF_{\text{Recall@50}}$		0,6	0.31	0,24

Bảng 4.2: So sánh hiệu năng mô hình đề xuất với các giải pháp đã có

Từ bảng kết quả thực nghiệm trong thí nghiệm này, có thể thấy mô hình cải tiến vượt trội hơn so với các giải pháp tốt nhất hiện nay trên cả hai tập dữ liệu với hầu hết các độ đo, khắc phục bài toán khởi động nguội, nâng cao hiệu năng tư vấn trên người dùng đã có tương tác và đảm bảo tính công bằng trong hệ tư vấn:

- Với bộ dữ liệu Movielens 1M, GIFT4Rec có hiệu năng tư vấn trên tập người dùng khởi động nguội tính theo Recall@50 tốt nhất, khiến độ đo GF vượt trội hơn DropoutNet. Tuy nhiên chất lượng tư vấn của mô GIFT4Rec lại tệ nhất trên tập người dùng cũ, kéo theo hệ lụy là độ đo GF và GF_2 thấp hơn so với mô hình cải tiến được đề xuất. Hai điều trên chứng minh tình trạng xấu gây ra bởi huấn luyện đồng thời biểu diễn tương tác đã được nêu tại chương 3, đồng thời chứng minh được mô hình khóa luận đề xuất đã giải quyết được các nhược điểm còn tồn đọng trong GIFT4Rec. DropoutNet có hiệu năng thể hiện qua độ đo GF rất tốt, nguyên nhân đến từ cách thức huấn luyện đặc biệt, học cách tái tạo lại đầu ra mô hình học biểu diễn tương tác. Thế nhưng, hiệu năng tư vấn của mô hình khóa luận khi được tính toán tất cả các độ đo khác cũng như GF dựa trên Recall@20 đều vượt trội hơn nhiều, nguyên nhân một phần đến từ cách thức huấn luyện không bị phụ thuộc nhiều vào mô hình học biểu diễn tương tác cũng như sức mạnh của các thành phần cải tiến khóa luận đề xuất. Mô hình MeLU có hầu hết các độ đo tệ nhất trong các mô hình trừ đại lượng GF . Điều này chứng minh tính đúng đắn của đại lượng GF_2 mô hình đề xuất, thể hiện được độ lớn của hiệu năng tư vấn các nhóm, khắc phục vấn đề hiệu năng tư vấn các nhóm càng thấp càng được đánh giá tốt từ đại lượng GF . Mô hình cải tiến khóa luận thể hiện kết quả trên tập dữ liệu Movielens 1M vượt trội trên hầu hết các độ đo so với các nhóm mô hình còn lại. Với những độ đo chưa nhận được kết quả tốt nhất thì mô hình đề xuất vẫn chiếm được vị trí số hai trong bảng so sánh.
- Trên bộ dữ liệu Douban, mô hình DropoutNet lại một lần nữa đạt độ đo GF tốt nhất do cách thức huấn luyện đặc biệt. Tuy nhiên, cũng chính do cách thức huấn luyện này, hiệu năng tư vấn của DropoutNet lại tệ khi được tính toán bằng các độ đo khác. Nguyên nhân đến từ hiệu năng của riêng mô hình biểu diễn tương tác trong DropoutNet không tốt. Mô hình GIFT4Rec tỏ ra vượt trội hơn nhiều so với DropoutNet, chứng minh vai trò lớn lao của thông tin bổ trợ tới việc đánh giá hành vi người dùng trong bộ dữ liệu này. Tuy nhiên, hiệu năng GIFT4Rec vẫn bị giới hạn do những nhược điểm còn tồn tại đã được nêu trong chương 3. Cuối cùng, mô hình khóa luận phát triển lại một lần nữa đã đạt được kết quả vượt trội

nhất trên bộ dữ liệu Douban tại hầu hết các độ đo, do khắc phục được các nhược điểm mà các mô hình còn lại mắc phải.

Thí nghiệm thứ hai để đánh giá mức độ quan trọng của từng thành phần cải tiến trong mô hình:

Dữ liệu	Tập người dùng / độ đo		Mô hình cải tiến (không huấn luyện riêng biệt)	Mô hình cải tiến (không cơ chế bảo vệ tham số quan trọng cho từng tác vụ)	Mô hình cải tiến
Movielens 1M	U_{test}	Recall@20	6.28	6.9	7.857
		Recall@50	12.867	14.063	15.927
	$U_{test_{cold-start}}$	Recall@20	13.547	13.72	13.806
		Recall@50	25.16	25.687	24.997
		$GF_{2Recall@20}$	9.916	10.31	10.83
		$GF_{2Recall@50}$	19.013	19.875	20.463
		$GF_{Recall@20}$	7.267	6.82	5.949
		$GF_{Recall@50}$	12.293	11.624	9.07
Douban	U_{test}	Recall@20	2.33	4.82	9.17
		Recall@50	7.33	15.27	16.18
	$U_{test_{cold-start}}$	Recall@20	7.18	4.95	7.41
		Recall@50	16.83	13.6	15.94
		$GF_{2Recall@20}$	4.76	4.89	8.29
		$GF_{2Recall@50}$	12.08	14.44	16.06
		$GF_{Recall@20}$	4.85	0.13	1.76
		$GF_{Recall@50}$	9.5	1.67	0.24

Bảng 4.3: So sánh hiệu năng mô hình đề xuất với các biến thể của nó đã bỏ đi một thành phần cải tiến

Với hiệu năng vượt trội của mô hình khóa luận trên hầu hết các độ đo so với các biến thể đã được lược bỏ một thành phần cải tiến, thí nghiệm đã chứng minh được vai trò quan trọng của từng thành phần cải tiến trong mô hình đề xuất của khóa luận.

Kết luận

Trong chương 4, khóa luận đã trình bày về phần cứng, phần mềm cũng như dữ liệu sử dụng. Cũng trong chương này, khóa luận đã trình bày và giải thích chi tiết kết

quả hai thí nghiệm với mục tiêu so sánh hiệu quả của mô hình đề xuất với các giải pháp tốt nhất hiện có và chứng minh từng thành phần cải tiến là không thể tách rời. Mô hình đề xuất trong cả hai thí nghiệm đều đạt hiệu năng tốt nhất trên hầu hết các độ đo.

Kết luận

Khóa luận này đã giới thiệu về hệ tư vấn cùng các thử thách lớn trong lớp bài toán này như vấn đề khởi động nguội hay vấn đề công bằng. Khóa luận nêu ra khuyến điểm các lớp giải pháp đang có giải quyết vấn đề người dùng khởi động nguội đang chưa để ý đến hiệu năng tư vấn của người dùng đã có tương tác khi kết hợp thêm thông tin hỗ trợ, dẫn đến hệ tư vấn thiên vị và thiếu công bằng.

Khóa luận đã trình bày cũng như phân tích điểm mạnh với điểm yếu của các giải pháp tốt nhất cho đến hiện nay giải quyết được vấn đề đặt ra là DropoutNet [10] và GIFT4Rec [17]. Lấy những nhược điểm của các giải pháp tốt nhất đã có làm động lực, khóa luận đề xuất cải tiến mô hình GIFT4Rec. Đây là mô hình được đề xuất trong bài báo “GIFT4Rec: An Effective Side Information Fusion Technique Apply to Graph Neural Network for Cold-Start Recommendation” của chính tác giả khóa luận thu được kết quả thực nghiệm tốt và được chấp thuận tại hội nghị ACIIDS 2023 có rank B.

Mô hình khóa luận đề xuất đã cải tiến cách thức huấn luyện trong mô hình DropoutNet và sử dụng trong mô hình GIFT4Rec giải quyết nhược điểm chí mạng của GIFT4Rec khi biểu diễn tương tác bị ảnh hưởng bởi thông tin hỗ trợ khi chúng không có ích trong việc đánh giá hành vi người dùng. Bên cạnh đó, khóa luận đề xuất giải pháp giúp giảm tình trạng quá khớp cũng như học mỗi tác vụ hiệu quả hơn, đồng thời giảm ảnh hưởng xấu từ sự khác nhau về phân phối giữa các tác vụ khi học đồng thời với mục tiêu chung là tăng hiệu năng và tính công bằng của hệ tư vấn.

Cuối cùng khóa luận đã tiến hành thực hiện hai thí nghiệm để chứng minh hiệu năng của mô hình đề xuất. Thí nghiệm đầu tiên đã cho thấy sự vượt trội của mô hình đề xuất trên hầu hết các độ đo so với các giải pháp tốt nhất đã có. Thí nghiệm cuối cùng đã chứng minh được vai trò và sự cần thiết của từng thành phần cải tiến trong mô hình khóa luận.

Tài liệu tham khảo

- [1]. Shoujin Wang, Qi Zhang, Liang Hu, Xiuzhen Zhang, Yan Wang, Charu Aggarwal, “Sequential/Session-based Recommendations: Challenges, Approaches, Applications and Opportunities”. SIGIR 2022: 3425–3428
- [2]. Suvash Sedhain, Scott Sanner, Darius Braziunas, Lexing Xie, and Jordan Christensen, “Social collaborative filtering for cold-start recommendations”. Recsys 2014: 345-348.
- [3]. Iman Barjasteh, Rana Forsati, Farzan Masrour, Abdol-Hossein Esfahanian, and Hayder Radha, “Cold-start item and user recommendation with decoupled completion and transduction”. Recsys 2015: 91–98
- [4]. Zeno Gantner, Lucas Drumond, Christoph Freudenthaler, Steffen Rendle, and Lars Schmidt-Thieme, “Learning Attribute-to-Feature Mappings for Cold-Start Recommendations”. ICDM 2010: 176–185
- [5]. Martin Saveski, Amin Mantrach, “Item cold-start recommendations: learning local collective embeddings”. Recsys 2014: 89–96
- [6]. Ajit P Singh and Geoffrey J Gordon, “Relational learning via collective matrix factorization”. KDD 2008: 650–658
- [7]. Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen, “Deep content-based music recommendation”. NIPS 2013: 2643–2651.
- [8]. Jingjing Li, Mengmeng Jing, Ke Lu, Lei Zhu, Yang Yang, and Zi Huang, “From Zero-Shot Learning to Cold-Start Recommendation”. AAAI 2019: 4189–4196.
- [9]. Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, Lexing Xie, and Darius Braziunas, “Low-rank linear cold-start recommendation from social data”. AAAI 2017: 1502–1508.
- [10]. Maksims Volkovs, Guangwei Yu, and Tomi Poutanen, “DropoutNet: Addressing cold start in recommender systems”. NIPS 2017: 4964–4973.
- [11]. Bin Li, Xingquan Zhu, Ruijiang Li, Chengqi Zhang, Xiangyang Xue, and Xindong Wu, “Cross-Domain Collaborative Filtering over Time”. ICJAI 2011: 2293–2298
- [12]. Liang Hu, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu, and Can Zhu, “Personalized recommendation via cross-domain triadic factorization”. WWW 2013: 595–606
- [13]. Pigi Kouki, Shobeir Fakhraei, James Foulds, Magdalini Eirinaki, and Lise Getoor, “Hyper: A flexible and extensible probabilistic framework for hybrid recommender systems”. RecSys 2015: 99–106.
- [14]. Hao Wang, Naiyan Wang, and Dit-Yan Yeung, “Collaborative deep learning for recommender systems”. KDD 2015: 1235–1244.
- [15]. Chong Wang and David M Blei, “Collaborative topic modeling for recommending scientific articles”. KDD 2011: 448–456 .

- [16]. David M Blei, Andrew Y Ng, and Michael I Jordan, “Latent dirichlet allocation. *Journal of machine Learning Research*”. JMLR 2003: 993–1022.
- [17]. Tran-Ngoc-Linh Nguyen, Chi-Dung Vu, Hoang-Ngan Le, Anh-Dung Hoang, Xuan Hieu Phan, Quang Thuy Ha, Hoang Quynh Le and Mai-Vu Tran “GIFT4Rec: An Effective Side Information Fusion Technique Apply to Graph Neural Network for Cold-Start Recommendation”. *ACIIDS 2023*: 334–345.
- [18]. Di Jin, Luzhi Wang, He Zhang, Yizhen Zheng, Weiping Ding, Feng Xia, Shirui Pan, “A Survey on Fairness-aware Recommender Systems”. *ACM Transactions on Information Systems 2023*: 1–43.
- [19]. Wan, Q., He, X., Wang, X., Wu, J., Guo, W., Tang, R., “Cross pairwise ranking for unbiased item recommendation”. *WWW 2022*: 2370–2378.
- [20]. Wang, Z., Xu, Q., Yang, Z., Cao, X., Huang, Q., “Implicit feedbacks are not always favorable: Iterative relabeled one-class collaborative filtering against noisy interactions”. *MM 2021*: 3070–3078.
- [21]. Montanari, M., Bernardis, C., Cremonesi, P., “On the impact of data sampling on hyper-parameter optimisation of recommendation algorithms”. *SAC 2022*: 1399–1402.
- [22]. Wang, J., Yang, Y., Wang, S., Hu, J., Wang, Q., “Contextand fairness-aware in-process crowdworker recommendation”. *ACM Transactions on Information Systems 2023*: 1–31.
- [23]. Zhu, Z., Kim, J., Nguyen, T., Fenton, A., Caverlee, J., “Fairness among new items in cold start recommender systems”. *SIGIR 2021*: 767–776.
- [24]. Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, Yongfeng Zhang, “Towards Personalized Fairness based on Causal Notion”. *SIGIR 2021*: 1054–1063.
- [25]. Hoyoep Lee, Jinbae Im, Seongwon Jang, Hyunsouk Cho, Sehee Chung, “MeLU: Meta-Learned User Preference Estimator for Cold-Start Recommendation”. *KDD 2019*: 1073–1082.