**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM TP. HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

**BÁO CÁO GIỮA KÌ**

**“CÁC KỸ THUẬT CHÍNH TRONG RS”**

**Học phần: 2021COMP131001** **– Hệ Tư Vấn Thông Tin**

Nhóm sinh viên thực hiện:

1. Nguyễn Hoàng Cát MSSV: 4501104021 2. Huỳnh Anh Dự MSSV: 4501104041 3. Nguyễn Văn Giàu MSSV: 4501104061 4. Lý Hoàng Long MSSV: 4501104129

5. Võ Thành Quang MSSV: 4501104189

Giảng viên hướng dẫn: Th.s Trần Thanh Nhã

**Thành phố Hồ Chí Minh, ngày 16 tháng 11 năm 2022**

**MỤC LỤC**

[MỞ ĐẦU 5](#_Toc119421817)

[Chương 1: CƠ SỞ LÝ THUYẾT 9](#_Toc119421818)

[1.1. Giới thiệu về Recommender Systems 9](#_Toc119421819)

[1.2. Các khái niệm chính 10](#_Toc119421820)

[Chương 2: CÁC KỸ THUẬT CHÍNH TRONG RS 10](#_Toc119421821)

[2.1. Kỹ thuật k láng giềng dựa trên người dùng/mục tin (User/Item kNNs) 11](#_Toc119421822)

[2.2. Kỹ thuật phân rã ma trận ( Matrix Factorization – MF) 11](#_Toc119421823)

[2.3. Các kỹ thuật không cá nhân hóa 11](#_Toc119421824)

[2.4. Các kỹ thuật nâng cao 11](#_Toc119421825)

[2.4.1. Kỹ thuật Phân rã ma trận thiên vị (Biased matrix factorization – BMF) 11](#_Toc119421826)

[2.4.2. Kỹ thuật Phân rã ma trận đa quan hệ (Multi-relational Matrix Factorization – MRMF) 11](#_Toc119421827)

[2.4.3. Gợi ý theo ngữ cảnh 12](#_Toc119421828)

[Chương 3. THIẾT KẾ WEBSITE THƯƠNG MẠI ĐIỆN TỬ 14](#_Toc119421829)

[3.1. Sơ đồ giao diện 14](#_Toc119421830)

[3.1.1 Sơ đồ giao diện phía Client 14](#_Toc119421831)

[3.1.2. Sơ đồ giao diện phía Admin 15](#_Toc119421832)

[3.2. Thiết kế giao diện 15](#_Toc119421833)

[3.2.1. Giao diện phía người dùng 15](#_Toc119421834)

[3.2.2 Giao diện phía admin 19](#_Toc119421835)

[Chương 4: TỔNG KẾT 22](#_Toc119421836)

[4.1. Kết quả đạt được 22](#_Toc119421837)

[4.2. Hướng phát triển 23](#_Toc119421838)

**LỜI CẢM ƠN**

Lời đầu tiên, nhóm chúng em xin gửi lời cảm ơn chân thành nhất đến thầy Trần Thanh Nhã. Trong quá trình học tập và tìm hiểu bộ môn Hệ tư vấn thông tin, chúng em đã nhận được sự quan tâm giúp đỡ và hướng dẫn rất tận tình từ tâm huyết của thầy. Thầy đã giúp chúng em tích lũy thêm nhiều kiến thức để có cái nhìn sâu sắc và hoàn thiện hơn trong cuộc sống. Thông qua bài báo cáo này nhóm chúng xin trình bày lại tất cả những gì mà mình đã đạt được trong thời gian qua.

Có lẽ kiến thức là vô hạn mà sự tiếp nhận kiến thức của bản thân mỗi người luôn tồn tại những hạn chế nhất định. Do đó, trong quá trình hình thành bài báo cáo chắc chắn sẽ không tránh khỏi những thiếu sót. Nhóm em rất mong nhận được những góp ý từ thầy để bài tiểu luận của nhóm được hoàn thiện hơn.

Kính chúc thầy sức khỏe, hạnh phúc và thành công trên con đường sự nghiệp giảng dạy.

**NHIỆM VỤ THÀNH VIÊN NHÓM**

|  |  |  |
| --- | --- | --- |
| **MSSV** | **Họ Tên** | **Phân Công Công Việc** |
| 4501104021 | Nguyễn Hoàng Cát | Tìm hiểu phần các kỹ thuật không cá nhân hóa |
| 4501104041 | Huỳnh Anh Dự | - Làm powerpoint  - Tổng hợp chỉnh sửa hoàn thành bài báo cáo - Nghiên cứu code  - Làm phần tổng quan về đề tài |
| 4501104061 | Nguyễn Văn Giàu | Tìm hiểu các kỹ thuật nâng cao |
| 4501104129 | Lý Hoàng Long | Tìm tài liệu phương pháp hệ gợi ý |
| 4501104189 | Võ Thành Quang | Tìm hiểu phần lọc cộng tác |

# MỞ ĐẦU

1. **Giới thiệu**

Recommender Systems (RSs) là những công cụ và kỹ thuật nhằm cung cấp đề xuất về một item cho người dùng. Những đề xuất này liên quan đến những quy trình ra quyết định khác nhau, chẳng hạn nên mua hàng gì, nghe nhạc gì hoặc đọc tin tức ở đâu. "Item" là thuật ngữ chung nhằm ám chỉ cái mà hệ thống đề xuất với người dùng. RSs chủ yếu hướng đến những cá nhân thiếu kinh nghiệm hoặc thẩm quyền để đánh giá tiềm năng của một số "item" mà chẳng hạn, một website có thể cung cấp. Trường hợp nếu là một hệ tư vấn sách thì nó sẽ hỗ trợ người dùng chọn lựa một cuốn sách để đọc. Rõ ràng, có khá nhiều lý do tại sao những website này lại muốn sử dụng RS:

* 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 RS. RS đề xuất càng chính xác với nhu cầu người dùng thì website càng bán được nhiều hàng.
* Thỏa mãn nhu cầu khách hàng: Một RS được thiết kế và cài đặt tốt sẽ cải thiện trải nghiệm người dùng. Người dùng sẽ thấy rằng hệ tư vấn này rất thú vị, chính xác, và cùng với một UI được thiết kế hợp lý, khách hàng sẽ muốn quay lại website.
* Nắm được nhu cầu khách hàng: Một chức năng quan trọng khác của RS là cho phép ta thu thập được phản hồi của khách hàng. 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 của RS.

Nghiên cứu và ứng dụng hệ gợi ý giúp nâng cao tính thông minh của ứng dụng, nó giải quyết được các câu hỏi bằng cách nào khai phá dữ liệu, xử lí và đưa ra thông tin cho người dùng một cách hiệu quả nhất trong trường hợp có hàng ngàn dữ liệu. Để trả lời câu hỏi đưa ra, yêu cầu đặt ra phải xây dựng được phương pháp phân loại, tìm kiếm, trích rút và đưa ra gợi ý kết quả cho người dùng.

Yêu cầu đặt ra: Xây dựng hệ thống gợi ý để gợi ý kết quả hoặc dữ liệu cho người dùng một cách hiệu quả với yêu cầu về thời gian, không gian

Để làm rõ yêu cầu đặt ra chúng ta sẽ thu hẹp phạm vi nghiên cứu bằng việc trả lời các câu hỏi:

* Hệ gợi ý dựa trên phương pháp lọc nào?
* Trường hợp có dữ liệu mới hoặc người dùng mới, hệ thống có đưa ra gợi ý hợp lý không?
* Bằng cách nào sử dụng thông tin về các đặc tính của dữ liệu kết hợp với phương pháp lọc?
* Hiệu quả của hệ thống gợi ý được đánh giá dựa trên các tiêu chí nào?

1. **Lý do chọn đề tài**

Người sử dụng các hệ thống thông tin, đặc biệt là các trang thương mại điện tử, thường gặp khó khăn trong việc tìm kiếm và lựa chọn các thông tin cần thiết và phù hợp để giải quyết một vấn đề nhất định, như việc chọn mua một máy ảnh kỹ thuật số phù hợp, hoặc việc lập kế hoạch cho một chuyến đi du lịch; bởi vì người sử dụng cỏ quá nhiều lựa chọn, nhưng không có đủ thời gian hoặc tri thức để tự đánh giá những lựa chọn này và đưa ra quyết định hợp lý. Bạn chọn mua cuốn sách nào giữa hàng triệu đầu sách được Amazon cung cấp, hay xem video trong số hàng trăm triệu video được tải lên Youtube. Bạn có thế tìm sự trợ giúp với công cụ tìm kiếm với những từ khóa cụ thể. Nhưng với trường hợp, bạn không biết rõ sự ưu tiên của mình, công cụ tìm kiếm khó có thể phát huy tác dụng, lúc này bạn cần đến sự giúp đỡ của hệ gợi ý. Các hệ thống gợi ý (Recommender System) là công cụ hỗ trợ ra quyết định, nhằm mục đích cung cấp cho người sử dụng những gợi ý về thông tin, sản phẩm, và dịch vụ phù hợp nhất với yêu cầu và sở thích riêng của từng người tại từng tình huống yêu cầu cụ thể.

Với sự trợ giúp của hệ gợi ý, bạn còn có thể tìm hiểu sở thích của mình và chia sẻ cho cộng đồng. Nếu bạn là fan hâm mộ ban nhạc The Beatles, một đề xuất đưa ra: bạn khó có thể bỏ qua những giai điệu của một nhóm nhạc người Anh khác là Radiohead. Radiohead rất nối tiếng trong ngành công nghiệp âm nhạc, và đặc biệt là ở Anh trong khoảng thập niên 90 với những ca khúc bất hủ “Creep”, Pablo Honey, hay The Bends... Khác với The Beatles, Radiohead không chơi thể loại nhạc pop đã làm nên tên tuổi của The Beatles, mà thay vào đó là dòng nhạc rock, hay punk rock. Vậy thì li do tại sao mà người hâm mộ Beatles, sẽ yêu thích Radiohead. Minh chứng cho điều này được thể hiện qua thông tin thu thập về hành vi trực tuyến của người dùng tại trang web Last.fm. Last.fm là một ứng cộng đồng âm nhạc, tập hợp thông tin về âm nhạc mà các thành viên lắng nghe, sau đó tổng hợp đưa ra các kết quả mang tính thống kê, giúp tìm hiểu thị hiếu âm nhạc và chia sẻ cho mọi người. Trang web giúp người dùng bằng cách gợi ý những nghệ sĩ, dòng nhạc đáng quan tâm. Cụ thể ở đây, những người nghe Beatles, thì những bài hát của những nghệ sĩ khác nào đáng để thưởng thức tiếp theo. Một danh sách những nghệ sĩ được Last.fm đưa ra, đứng đầu trong danh sách là Bob Dylan - Một nghệ sĩ nổi tiếng cùng thời với The Beatles, và tiếp theo trong danh sách là Radio head. Theo Last.fm, người hâm mộ Beatles nghe nhiều của Radiohead, và ngược lại.

1. **Mục tiêu và nhiệm vụ nghiên cứu**

Nghiên cứu hệ thống gợi ý: Cách thức hoạt động, mô hình, các phương pháp, tầm quan trọng của hệ gợi ý.

Lựa chọn, nghiên cứu, và cài đặt phương pháp gợi ý phù hợp cho bài toán đặt ra: Xây dựng hệ gợi ý giúp người dùng lựa chọn những dữ liệu phù hợp

1. **Đối tượng và phạm vi nghiên cứu**

Đề tài nghiên cứu về hệ gợi ý và ứng dụng triển khai trong website gợi ý dữ liệu cho người dùng.

Đối tượng gợi ý: Các dữ liệu được thu thập lưu trong cơ sở dữ liệu

Đối tượng được gợi ý:

* Người dùng tìm kiếm trên hệ thống và hành vi được lưu lại trên hệ thống
* Người dùng không có trên hệ thống, ghé thăm hệ thống để tìm kiếm lựa phim

Nội dung gợi ý: là tập các dữ liệu được hệ thống cho là phù hợp nhất với người dùng trong một ngữ cảnh xác định

# Chương 1: CƠ SỞ LÝ THUYẾT

## 1.1. Giới thiệu về Recommender Systems

Hệ thống gợi ý (Recommender Systems - RS) là một dạng của hệ thống lọc thông tin (information filtering), nó được sử dụng để dự đoán sở thích (preferences) hay xếp hạng (rating) mà người dùng có thể dành cho một mục thông tin (item) nào đó mà họ chưa xem xét tới trong quá khứ (item có thể là bài hát, bộ phim, đoạn video clip, sách, bài báo,..).

Ví dụ, trong hệ thống bán hàng trực tuyến (chẳng hạn như Amazon), nhằm tối ưu hóa khả năng mua sắm của khách hàng (user), người ta quan tâm đ ến việc những khách hàng nào đã ‘yêu thích’ những sản phẩm (item) nào bằng cách dựa vào dữ liệu quá khứ của họ (dữ liệu này có thể là xếp hạng mà người dùng đã bình chọn trên sản phẩm, thời gian duyệt (browse) trên sản phẩm, số lần click chuột trên sản phẩm,..) từ đó hệ thống sẽ dự đoán được người dùng có thể thích sả n phẩm nào và đưa ra những gợi ý phù hợp cho họ. Hình 1 là một ví dụ minh họa cho hệ thống gợi ý bán hàng của Amazon.

Ngoài lĩnh vực thương mại điện tử như đã thấy ở ví dụ trên, hiện tại RS cũng được ứng dụng khá thành công trong nhiều lĩnh vực khác như trong giải trí: gợi ý bài hát cho người nghe (ví dụ, hệ thống của LastFM - www.last.fm), gợi ý phim ảnh (ví dụ, hệ thống của Netflix - www.netflix.com), gợi ý các video clip (ví dụ, hệ thống của YouTube - www.youtube.com); trong giáo dục và đào tạo (gợi ý nguồn tài nguyên học tập như sách, bài báo, địa chỉ web,… cho người học); trong các hệ thống trợ giảng thông minh (như dự đoán năng lực học tập của người học)

Graphical user interface, text, application, email

Description automatically generated

Hình 1. Hệ thống gợi ý sản phẩm của Amazon

## 1.2. Các khái niệm chính

Trong RS, thông thường người ta quan tâm đến ba thông tin chính là **người dùng** (user), **mục tin** (item, item có thể là sản phẩm, bộ phim, bài hát, bài báo,.. tùy hệ thống), và **phản hồi** (feedback) của người dùng trên mục tin đó (thư ờng là các **xếp hạng**/đánh giá – rating biểu diễn mức đ ộ thích/quan tâm của họ). Các thông tin nàyđược biểu diễn thông qua một ma trận như trong Hình 2. Ở đó, mỗi dòng là một user, mỗi cột là một item, và mỗi ô là một giá trị phản hồi (ví dụ, xếp hạng) biểu diễn “mức độ thích” của user trên item tương ứng. Các ô có giá trị là những item mà các user đã xếp hạng trong quá khứ. Những ô trống là những item chưa được xếp hạng (điều đáng lưu ý là mỗi user chỉ xếp hạng cho một vài item trong quá khứ, do vậy có rất nhiều ô trống trong ma trận này – còn gọi là ma trận thưa – sparse matrix).

A picture containing text, crossword puzzle

Description automatically generated

Hình 2. Ma trận biểu diễn dữ liệu trong RS (user-item-rating matrix)

Nhiệm vụ chính của RS là dựa vào các ô đã có giá tr ị trong ma trận trên (dữ liệu thu được từ quá khứ), thông qua mô hình đã đư ợc xây dựng, RS dự đoán các ô còn trống (của user hiện hành), sau đó sắp xếp kết quả dự đoán (ví dụ, từ cao xuống thấp) và chọn ra Top-N items theo thứ tự, từ đó gợi ý chúng cho người dùng.

# Chương 2: CÁC KỸ THUẬT CHÍNH TRONG RS

Hiện tại, trong RS có rất nhiều giải thuật được đề xuất, tuy nhiên có thể gom chúng vào trong các nhóm chính

* Nhóm giải thuật **lọc cộng tác** (Collaborative Filtering): trong nhóm này, các giải thuật chủ yếu sử dụng:
* Phương pháp láng giềng (Neighborhood-based, còn gọi là Memory-based), trong đó hoặc là dựa trên dữ liệu quá khứ của người dùng “tương tự - similarity” (user-based approach), hoặc là dựa trên dữ liệu quá khứ của những item “tương tự” (item-based approach)
* Dựa trên mô hình (Model-based): Nhóm này liên quan đ ến việc xây dựng các mô hình dự đoán dựa trên dữ liệu thu thập được trong quá khứ. Như mô hình Bayesian, các mô hình nhân tố tiềm ẩn (latent factor models): trong đó kỹ thuật phân rã ma trận (matrix factorization) là một điển hình
* Nhóm giải thuật **lọc trên nội dung** (Content-based Filtering): Gợi ý các item dựa vào hồ sơ (profiles) của người dùng hoặc dựa vào nội dung/thuộc tính (attributes) của những item tương tự như item mà người dùng đã chọn trong quá khứ.
* Nhóm lai ghép: Kết hợp cả 2 cách trên
* Nhóm kỹ thuật không cá nhân hóa ( non-personalization)

Một trong những khuyết điểm của phương pháp lọc theo nội dung là khó khăn trong việc thu thập thông tin, trong khi phần lớn các mô hình dựa trên lọc cộng tác chỉ cần 3 thông tin (user id, item id, feedback) là có thể hoạt động tốt. Do vậy khuynh hướng hiện nay đa phần các nhà nghiên cứu thiên về nhóm lọc cộng tác.

Sau đây chúng tôi sẽ trình bày tóm tắt các kỹ thuật thông dụng và nổi trội (state-of-the-art) trong hệ thống gợi ý.

## 2.1. Kỹ thuật k láng giềng dựa trên người dùng/mục tin (User/Item kNNs)

Có hai cách tiếp cận của lọc cộng tác theo mô hình K láng giềng là dựa trên người dùng (User\_kNN) – dự đoán dựa trên sự tương tự giữa các người dùng và dựa trên mục tin (Item\_kNN) – dự đoán dựa trên sự tương tự giữa các mục tin.

Kỹ thuật k láng giềng dựa trên người dùng (User\_kNN) xác định độ tương tự giữa hai người dùng thông qua việc so sánh các đánh giá của họ trên cùng sản phẩm, sau đó dự đoán xếp hạng trên sản phẩm i bởi người dùng u, thông qua các xếp hạng của những người dùng tương tự với người dùng u. Độ tương tự giữa người dùng u và người dùng u' có thể được tính theo Cosine hoặc Pearson như sau:

Text, letter

Description automatically generated

Trong đó:

rui và ru’i là đánh giá của người dùng u và u’ trên item i tương ứng

Tương tự, kỹ thuật k láng giềng dựa trên mục tin (Item\_kNN) cũng xác định độ tương tự dựa trên các mục tin bằng phương pháp Cosine hoặc Pearson như sau:

Text, letter

Description automatically generated

Trong đó:

*Uii'* là tập các người dùng có đánh trên cả hai mục tin i và i’

Sau khi tính toán độ tương tự giữa các người dùng hay giữa các sản phẩm, đánh giá của người dùng u trên sản phẩm i được dự đoán theo các công thức như bên dưới:

* Với phương pháp User\_kNN, xếp hạng (đánh giá/ feedback) của người dùng u trên sản phẩm i được dự đoán qua công thức

A picture containing text

Description automatically generated

Trong đó *Sim(u,u')* là độ tương tự giữa người dùng u và u' được xác đ ịnh bằng phương pháp Cosine hoặc Pearson

* Với phương pháp Item\_kNN, xếp hạng (đánh giá/ feedback) của người dùng u trên sản phẩm i được dự đoán qua công thức

Text

Description automatically generated

Trong đó, *Sim(i, i')* độ tương tự giữa mục tin i và i’ được xác định bằng phương pháp Cosine hoặc Pearson như đã trình bày; *Ki* là số item có độ lân cận gần item i (k láng giềng của i).

## 2.2. Kỹ thuật phân rã ma trận ( Matrix Factorization – MF)

Kỹ thuật phân rã ma trận là việc chia một ma trận lớn **X** thành hai ma trận có kích thước nhỏ hơn **W** và **H**, sao cho ta có thể xây dựng lại **X** từ hai ma trận nhỏ hơn này càng chính xác càng tốt

Trong đó, **W** ∈ **ℜ** là một ma trận mà mỗi dòng *u* là một véc tơ bao gồm K nhân

tố tiềm ẩn (latent factors) mô tả người dùng *u;* và **H** ∈ **ℜ** là một ma trận mà mỗi dòng *i* là một véc tơ bao gồm K nhân tố tiềm ẩn mô tả cho item *i*

Diagram

Description automatically generated with low confidence

Hình 3. Minh họa kỹ thuật phân rã ma trận

Gọi *wuk* và *hik* là các phần tử tương ứng của hai ma trận **W** và **H**, khi đó xếp hạng của người dùng *u* trên mục tin *i* được dự đoán bởi công thức:

Text

Description automatically generated with medium confidence

Như vậy vấn đề then chốt của kỹ thuật MF là làm sao để tìm được giá trị của hai tham số **W** và **H.** Hai tham số này có được bằng cách tối ưu hóa hàm mục tiêu (objective function). Hàm mục tiêu thông dụng nhất trong RS tương tự như hàm lỗi RMSE, biểu diễn như sau:

A picture containing text, watch, gauge

Description automatically generated

## 2.3. Các kỹ thuật không cá nhân hóa

Trong nhóm kỹ thuật này, do chúng khá đơn gi ản, dễ cài đ ặt nên nên thường được các website/hệ thống tích hợp vào, gồm cả các website thương mại, website tin tức, hay giải trí. Chẳng hạn như trong các hệ thống bán hàng trực tuyến, người ta thường gợi ý các sản phẩm được xem/mua/bình luận/.. nhiều nhất; gợi ý các sản phẩm mới nhất; gợi ý các sản phẩm cùng loại/ cùng nhà sản xuất/..; gợi ý các sản phẩm được mua/chọn cùng nhau. Một ví dụ khá điển hình là thông qua luật kết hợp (như Apriori), Amazon đã áp dụng khá thành công để tìm ra các sản phẩm hay được mua cùng nhau như minh họa trong Hình 4.

Tuy vậy, bất lợi của các phương pháp này là không cá nhân hóa cho từng người dùng, nghĩa là tất cả các user đều được gợi ý giống nhau khi chọn cùng sản phẩm.

Graphical user interface

Description automatically generated

Hình 4. Gợi ý sản phẩm thường được mua cùng nhau

## 2.4. Các kỹ thuật nâng cao

### 2.4.1. Kỹ thuật Phân rã ma trận thiên vị (Biased matrix factorization – BMF)

Kỹ thuật Phân rã ma trận thiên vị BMF là một biến thể của kỹ thuật Matrix Factorization (MF) cộng thêm các giá trị thiên vị/lệch (biases).

Với kỹ thuật BMF, xếp hạng của người dùng u cho mục tin i được dự đoán như sau:

Diagram

Description automatically generated with medium confidence

Trong đó, µ là giá trị trung bình toàn cục, được xác định từ tập dữ liệu huấn luyện qua công thức

Text

Description automatically generated

Text

Description automatically generatedGiá trị bu là độ thiên vị (bias) của người dùng u, xác định bởi:

Giá trị bi là độ thiên vị (bias) của item i, xác định bởi:

Text

Description automatically generated

Hàm mục tiêu của BMF có dạng:

A picture containing text

Description automatically generated

Tương tự như kỹ thuật MF, sau khi tối ưu hóa hàm mục tiêu, ta cũng xác định được các tham số tương ứng và từ đó có thể dễ dàng dự đoán kết quả.

Thực nghiệm đã cho th ấy, trong khá nhiều trường hợp, BMF cho kết quả tốt hơn hẵn so với kỹ thuật MF

### 2.4.2. Kỹ thuật Phân rã ma trận đa quan hệ (Multi-relational Matrix Factorization – MRMF)

Kỹ thuật phân rã ma trận đa quan hệ (Multi-relational Matrix Factorization – MRMF) là kỹ thuật mở rộng từ kỹ thuật MF với ý tưởng là tận dụng tất cả các mối quan hệ có thể có trong cơ sở dữ liệu của hệ thống.

Nếu như trong MF, ta chỉ sử dụng 1 ma trận (tương ứng với 1 quan hệ trong sơ đồ ERD, ví dụ như “Entity1-Relation1-Entity2” trong Hình 5 thì trong MRMF ta sẽ sử dụng tất cả các ma trận tương ứng với các thực thể và các quan hệ có thể có, như minh họa trong Hình 5.

Diagram

Description automatically generated

Hình 5. Biểu diễn dữ liệu trong MRMF

Hàm mục tiêu của MRMF được viết dưới dạng tổng quát cho N thực thể và M quan hệ như sau:

Text

Description automatically generated with low confidence

Tương tự trong MF, sử dụng stochastic gradient descent, ta cũng xây dựng được mô hình MRMF như trong thủ tục bên dưới. Trong đó

Text, letter

Description automatically generated

### 2.4.3. Gợi ý theo ngữ cảnh

Trong RS hai chiều như đã trình bày, người ta thường biểu diễn dữ liệu vào ma trận user-item do vậy chỉ quan tâm đến **người dùng** và **mục tin** là chính chứ không quan tâm đến các **thông tin ngữ cảnh** (context) bên ngoài có tác động đến quyết định của người dùng hay không. “*Thông tin ngữ cảnh là những thông tin có thể mô tả được hoàn cảnh của một thực thể*” . Ví dụ: thời gian, nơi chốn, thờitiết, tâm trạng… là những thông tin ngữ cảnh, chúng có thể ảnh hưởng đến đánh giá của người dùng đối với các item, từ đó dẫn đến sự ảnh hưởng của những gợi ý trong RS.

Với hệ thống gợi ý theo ngữ cảnh (đa chiều), hàm dự đoán được bổ sung thêm thông tin ngữ cảnh (Context) và trở thành:



Ví dụ, trong hệ thống gợi ý du l ịch, user là khách du lịch, item là những địa điểm du lịch, và ngữ cảnh có thể là *thời gian* (buổi sáng/tối, cuối tuần, mùa, lễ, tết,..), *bạn đ ồng hành* (đi một mình, đi v ới bạn trai/bạn gái, đi v ới gia đình, tr ẻ nhỏ…)…Những đánh giá cho m ột đi ểm du lịch bởi một người dùng có thể sẽ phụ thuộc vào những ngữ cảnh đó. Ví dụ vào buổi tối mùa xuân đi cùng gia đình thì địa điểm ABC là hấp dẫn nhất, nhưng khi đi cùng bạn gái thì có thể địa điểm đó không phù hợp nữa.

Các phương pháp tiếp cận trong hệ thống gợi ý đa chiều có thể được tóm tắt như sau:

* Xử lý ngữ cảnh đầu vào (Contextual pre-filtering): Trong tiếp cận này, thông tin về ngữ cảnh được sử dụng để lựa chọn hoặc xây dựng các thiết lập có liên quan đến dữ liệu (xếp hạng). Sau khi lọc thông tin ngữ cảnh, hệ thống đa chiều sẽ trở thành hệ thống hai chiều và ta có thể sử dụng bất kỳ phương pháp gợi ý truyền thống nào trên các dữ liệu đã chọn
* Xử lý ngữ cảnh đầu ra (Contextual post-filtering): Trong tiếp cận này, ngữ cảnh thông tin ban đầu bị bỏ qua, và xếp hạng được dự đoán bằng cách sử dụng bất kỳ phương pháp gợi ý truyền thống trên toàn bộ dữ liệu. Sau đó, các gợi ý đư ợc điều chỉnh cho mỗi người sử dụng bằng cách sử dụng các thông tin theo ngữ cảnh
* Mô hình hóa ngữ cảnh (Contextual modeling): Phương pháp này xây dựng một mô hình dự đoán. Ví dụ, mô hình này có thể đưa ra xác suất một người dùng cụ thể chọn một sản phẩm cụ thể trong một bối cảnh cụ thể là bao nhiêu. Sau đó xác suất này được dùng để điều chỉnh và tìm ra gợi ý phù hợp

Để xây dựng hệ thống gợi ý theo ngữ cảnh thì có thể xử dụng phương pháp như hình minh họa trong *Hình 6.* Trước tiền xử lý thông tin ngữ cảnh đầu vào, kế đến dùng kỹ thuật phân rã ma trận để dự đoán kết quả, sau đó xử lý thông tin ngữ cảnh đầu ra

Text

Description automatically generated

Hình 6. Quy trình gợi ý theo ngữ cảnh

**Xử lý thông tin ngữ cảnh đầu vào:** Phương pháp này dùng những đi ều kiện ngữ cảnh hiện tại được cung cấp trực tiếp từ người dùng (sở thích, tâm trạng…) hoặc từ môi trường (thời gian, thời tiết, vị trí hiện tại…) sau đó truy v ấn, tìm kiếm những nguồn tài nguyên thích hợp nhất để đưa ra gợi ý. Dữ liệu đầu vào là một tập bao gồm: người dùng, đối tượng, ngữ cảnh và những đánh giá tương ứng. Sau đó dùng điều kiện ngữ cảnh hiện tại của người dùng cung cấp để lọc ra tập dữ liệu có liên quan đến ngữ cảnh đó. Sau khi l ọc ta có thể bỏ qua thông tin ngữ cảnh, như vậy bài toán gợi ý đa chiều lúc này trở về dạng bài toán gợi ý hai chiều như minh họa trong Hình 7. Từ đây, ta có thể áp dụng bất kỳ phương pháp gợi ý hai chi ều nào đ ể dự đoán các item cho người dùng. Lợi ích của việc này là có thể tái sử dụng tất cả những phương pháp gợi ý hai chiều sau khi chiều ngữ cảnh được thu giảm

Diagram, engineering drawing

Description automatically generated

Hình 7. Sử dụng phương pháp Pre-filtering

**Xử lý thông tin ngữ cảnh đầu ra:** Trong phương pháp này, thông tin ngữ cảnh được sử dụng sau khi đã có lời gợi ý. Đầu vào của phương pháp này vẫn là tập dữ liệu bao gồm cả thông tin ngữ cảnh. Tuy nhiên, ta giả định các thông tin ngữ cảnh chưa ảnh hưởng đến các đánh giá của người dùng lên các đối tượng. Sau đó dùng những kỹ thuật gợi ý hai chiều để tiến hành dự đoán các đánh giá và gợi ý trên toàn bộ tập dữ liệu ban đầu. Sau khi có kết quả gợi ý cho người dùng, ta sẽ lọc ra những lời gợi ý phù hợp với thông tin ngữ cảnh ban đầu rồi mới chuyển đến người dùng.

Thủ tục tổng quát cho xây dựng hệ thống gợi ý theo ngữ cảnh kết hợp xử lý ngữ cảnh đầu vào, dùng mô hình phân rã ma trận, và xử lý ngữ cảnh đầu ra được minh họa trong thủ tục **ContextAware-MF**. Các siêu tham số (hyper-parameters): *Số lần lặp (*Iter), *Số nhân tố tiềm ẩn* K, *tốc độ học* β, và hệ số chính tắc hóa λ (*regularization)* nhằm tránh học vẹt (over-fitting).

# Chương 3. CÁC BÀI NGHIÊN CỨU TRONG VÀ NGOÀI NƯỚC

1. Bobadilla J., Ortega F., Hernando A., Gutiérrez H. 2013. Recommender systems survey. Knowledge-Based Systems. 46 (2013) 109–132. Elsevier

Link: <https://www.sciencedirect.com/science/article/abs/pii/S0950705113001044?via%3Dihub>

2. Lư Chân Thiện và **Nguyễn Thái Nghe**. 2015. MỘT TIẾP CẬN TRONG XÂY DỰNG HỆ THỐNG GỢI Ý THEO NGỮ CẢNH. Kỷ yếu Hội nghị quốc gia lần thứ VIII "Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin" (FAIR 2015)

Link: <https://www.researchgate.net/publication/292281409_MOT_TIEP_CAN_TRONG_XAY_DUNG_HE_THONG_GOI_Y_THEO_NGU_CANH>

3. Nguyễn Hùng Dũng, **Nguyễn Thái Nghe**. 2014. Hệ thống gợi ý sản phẩm trong bán hàng trực tuyến sử dụng kỹ thuật lọc cộng tác. Tạp chí Khoa học Trường Đại học Cần Thơ, số 31a (2014), trang 36-51. ISSN: 1859-2333

Link:

<https://ctujsvn.ctu.edu.vn/index.php/ctujsvn/article/view/1394/2425>

1. Recommender Systems Prem Melville and Vikas Sindhwani IBM T.J. Watson Research Center, Yorktown Heights, NY 10598

Link: <https://www.ime.usp.br/~jstern/miscellanea/seminario/Melville1.pdf>

1. Recommender Systems in E-Commerce

Link: <https://dl.acm.org/doi/pdf/10.1145/336992.337035>

# Chương 4: TỔNG KẾT