

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



**Đặng Văn Tuấn**

# **XÂY DỰNG HỆ THỐNG GỢI Ý DU LỊCH THEO NGŨ CẢNH**

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

**HÀ NỘI - 2022**

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



**Đặng Văn Tuấn**

# **XÂY DỰNG HỆ THỐNG GỢI Ý DU LỊCH THEO NGŨ CẢNH**

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

**Cán bộ hướng dẫn: TS. Vũ Thị Hồng Nhạn**

**HÀ NỘI - 2022**

## LỜI CẢM ƠN

Lời đầu tiên, em xin được bày tỏ lòng biết ơn sâu sắc đến tất cả thầy cô trường Đại học Công Nghệ - Đại học Quốc Gia Hà Nội đã tận tình truyền đạt kiến thức quý báu của mình cho em trong những năm học tập ở trường. Em luôn tự hào vì được học tập và phát triển bản thân trong môi trường đầy đủ cơ sở vật chất cùng với đội ngũ giảng viên giàu kinh nghiệm và đầy nhiệt huyết giúp đỡ mọi sinh viên.

Đặc biệt em xin gửi lời cảm ơn chân thành tới TS. Vũ Thị Hồng Nhạn đã tận tình giảng dạy, giúp em định hướng và có những đóng góp quý báu giúp em có thể hoàn thành khóa luận này một cách tốt nhất.

Cuối cùng em xin được gửi lời cảm ơn đến gia đình, tập thể lớp K63J đã cùng nhau gắn bó và giúp đỡ nhau trong suốt quá trình học tập tại trường

Em xin chân thành cảm ơn!

# XÂY DỰNG HỆ THỐNG GỢI Ý DU LỊCH THEO NGŨ CẢNH

## Tóm tắt khóa luận

**Tóm tắt:** Trong cuộc sống hiện nay với sự phát triển về công nghệ đời sống con người xã hội ngày càng phát triển. Cùng với sự tiến bộ của xã hội thì nhu cầu về du lịch cũng ngày càng tăng. Trong những năm gần đây du lịch trở thành ngành dịch vụ mang lại lợi nhuận rất lớn cho các quốc gia. Mặc dù có rất nhiều địa điểm du lịch thú vị, nhưng du khách vẫn rất đau đầu cho việc lựa chọn địa điểm du lịch phù hợp với từng nhu cầu của mỗi người. Tuy các thông tin về địa điểm du lịch đã có rất nhiều trên các trang mạng nhưng họ vẫn không thể chọn lọc được thông tin và xác định được địa điểm phù hợp với mình. Dựa trên suy nghĩ đó tôi đã xây dựng một hệ thống đáp ứng cho nhu cầu tiên quyết của du khách. Theo đó khóa luận của tôi tiến hành xây dựng một **Hệ thống gợi ý du lịch theo ngũ cảnh**. Hệ thống được xây dựng và phát triển bởi Framework **React**, ngôn ngữ **Nodejs** cùng với **Google Map** để đưa ra gợi ý các địa điểm phù hợp nhất với sở thích, hoàn cảnh của từng cá nhân trong một thời gian ngắn ngày.

**Từ khóa:** Hệ thống gợi ý du lịch theo ngũ cảnh, Reactjs, Nodejs

## **LỜI CAM ĐOAN**

Tôi xin cam đoan bài khóa luận này đều do tôi thực hiện dưới sự hướng dẫn và giám sát của TS. Vũ Thị Hồng Nhạn. Tất cả các tài liệu liên quan trong quá trình nghiên cứu đều được ghi rõ nguồn gốc ở cuối luận văn. Tôi xin cam kết không sao chép tài liệu, công trình nghiên cứu của người khác. Tất cả kết quả thực nghiệm đều được lấy ra từ chương trình mà tôi phát triển. Nếu sai sự thật tôi xin chịu mọi trách nhiệm của trường.

Hà Nội, ngày 09 tháng 06 năm 2022

Tác giả khóa luận

Đặng Văn Tuấn

# MỤC LỤC

Tóm tắt khóa luận .....	1
LỜI CAM ĐOAN .....	2
DANH MỤC CHỮ VIẾT TẮT .....	5
DANH MỤC CÁC BẢNG .....	6
DANH MỤC CÁCH HÌNH.....	7
Chương 1: Lời mở đầu.....	9
Chương 2: Cơ sở lý luận.....	11
2.1 Giới thiệu.....	11
2.1.1. Đề tài .....	11
2.1.2. Mục tiêu .....	11
2.1.3. Phạm vi .....	11
2.1.4. Phương pháp tiếp cận.....	12
2.2. Tổng quan về hệ thống gợi ý .....	12
2.2.1. Hệ thống gợi ý truyền thống .....	12
2.2.2. Hệ thống gợi ý trên ngữ cảnh.....	12
2.2.3. Bài toán gợi ý trên ngữ cảnh .....	13
2.2.4. Cách thức hoạt động của hệ thống gợi ý.....	15
2.2.5. Tính ứng dụng của hệ thống gợi ý .....	17
Chương 3: Thực trạng.....	19
3.1. Phương pháp lọc ngữ cảnh trước.....	19
3.2. Phương pháp lọc ngữ cảnh sau.....	32
3.3. Phân chia gợi ý theo ngữ cảnh .....	34
Chương 4: Giải pháp – Xây dựng ứng dụng.....	37
4.1. Tập dữ liệu.....	37
4.1.1. Giới thiệu về tập dữ liệu.....	37
4.1.2. Phương pháp thu thập dữ liệu .....	39

4.2. Tổng quan về hệ thống .....	40
4.2.1. Các thành phần chính của hệ thống .....	40
4.2.2. Mô hình triển khai.....	40
4.3. Yêu cầu hệ thống .....	41
4.3.1. Phân tích yêu cầu .....	41
4.3.2. Xác định tác nhân.....	41
4.3.3. Bảng thuật ngữ.....	42
4.3.4. Biểu đồ ca sử dụng – Usecase Diagram.....	42
4.3.5. Kịch bản .....	43
4.4. Thiết kế hệ thống.....	50
4.4.1. Sơ đồ tuần tự .....	50
4.4.2. Biểu đồ thực thể thiết kế.....	53
4.4.3. Lược đồ cơ sở dữ liệu .....	53
4.5. Kết quả.....	54
Chương 5 Kết luận .....	59
TÀI LIỆU THAM KHẢO .....	60

## DANH MỤC CHỮ VIẾT TẮT

KÝ HIỆU	DIỄN GIẢI
CF	Collaborative Filtering (Lọc cộng tác)
NBCF	Neighborhood-based Collaborative Filtering (Lọc cộng tác dựa trên láng giềng)
IBF	Item-base Filter (Lọc dựa trên độ giống nhau của sản phẩm – địa điểm)
IR	Information retrieval (Thu thập thông tin)
IF	Information filtering (Lọc thông tin)
MF	Matrix Factorization (Phân rã ma trận)
UT	Utility Matrix



## DANH MỤC CÁC BẢNG

Bảng 2.1 Bảng dữ liệu đánh giá chứa thông tin ngữ cảnh .....	14
Bảng 3.1 Bảng đánh giá người dùng với địa điểm du lịch trong các điều kiện ngữ cảnh .....	21
Bảng 3.2 Bảng dữ liệu của một người dùng sau khi sử dụng phương pháp lọc trước .	21
Bảng 3.3 Dữ liệu đánh giá của người dùng sau khi lọc ngữ cảnh.....	23
Bảng 3.4 Bảng dữ liệu User-Item sau khi lọc ngữ cảnh.....	23
Bảng 3.5 Kết quả kiểm nghiệm của hệ thống gợi ý .....	31
Bảng 3.6 Bảng user-item sau khi chiếu lên mặt phẳng .....	33
Bảng 3.7 Bảng dữ liệu đánh giá địa điểm trong ngữ cảnh .....	35
Bảng 4.1 Bảng thuật ngữ .....	42
Bảng 4.2 Kịch bản người dùng đăng nhập .....	44
Bảng 4.3 Kịch bản người dùng đăng ký .....	45
Bảng 4.4 Kịch bản người dùng tra cứu hệ thống gợi ý .....	46
Bảng 4.5 Kịch bản người dùng xem kinh nghiệm người dùng .....	47
Bảng 4.6 Kịch bản người dùng thêm, xem bộ sưu tập .....	48
Bảng 4.7 Kịch bản người dùng đánh giá .....	49

## DANH MỤC CÁCH HÌNH

Hình 2.1 Minh họa cho không gian gợi ý ba chiều .....	14
Hình 2.2.Các thành phần chính của hệ thống gợi ý.....	15
Hình 2.3 Quá trình hoạt động chi tiết của hệ thống .....	16
Hình 3.1 Phương pháp lọc ngữ cảnh trước.....	19
Hình 3.2 Các mảnh dữ liệu ngữ cảnh .....	24
Hình 3.3 Tiến trình lọc cộng tác .....	25
Hình 3.4 Utility matrix Y được phân tích thành tích hai ma trận low rank X và W ....	27
Hình 3.5 Kết quả kiểm nghiệm.....	32
Hình 3.6 Phương pháp lọc ngữ cảnh sau.....	32
Hình 3.7 Quá trình hoạt động của hệ thống gợi ý truyền thống.....	34
Hình 4.1 Bộ dữ liệu Tourpedia, kết hợp với Google.....	38
Hình 4.2 Dữ liệu địa điểm thu thập được bằng google map api.....	39
Hình 4.3 Dạng JSON đánh giá của người dùng .....	40
Hình 4.4 Tổng quan về hệ thống .....	41
Hình 4.5 Use Case tổng quát của hệ thống.....	43
Hình 4.6 Biểu đồ tuần tự - Người dùng đăng nhập .....	50
Hình 4.7 Biểu đồ tuần tự - Người dùng đăng ký .....	50
Hình 4.8 Biểu đồ tuần tự - Người dùng tra cứu hệ thống gợi ý du lịch .....	51
Hình 4.9 Biểu đồ tuần tự - Người dùng xem kinh nghiệm du lịch.....	51
Hình 4.10 Biểu đồ tuần tự - Người dùng thêm, xem bộ sưu tập .....	52
Hình 4.11 Biểu đồ tuần tự - Người dùng đánh giá địa điểm .....	52
Hình 4.12 Biểu đồ thực thể.....	53
Hình 4.13 Lược đồ cơ sở dữ liệu .....	53
Hình 4.14 Giao diện màn hình đăng nhập.....	54
Hình 4.15 Giao diện màn hình đăng ký.....	54
Hình 4.16 Giao diện màn hình trang chủ .....	55
Hình 4.17 Giao diện màn hình kinh nghiệm .....	55

Hình 4.18 Giao diện màn hình bộ sưu tập.....	56
Hình 4.19 Màn hình gợi ý ban đầu .....	56
Hình 4.20 Màn hình gợi ý khi nhập vị trí xuất phát và kết quả.....	57
Hình 4.21 Màn hình gợi ý với điểm xuất phát là vị trí hiện tại .....	57
Hình 4.22 Màn hình kết quả với điểm xuất phát là vị trí hiện tại .....	58
Hình 4.23 Màn hình đánh giá địa điểm .....	58

## Chương 1: Lời mở đầu

Trước cuộc cách mạng công nghiệp 4.0, dẫn tới sự phát triển không ngừng của xã hội, theo đó là các ngành nghề, dịch vụ trong và ngoài nước cũng liên tục bùng nổ, tạo ra sự tăng trưởng mạnh mẽ về kinh tế, nhu cầu đời sống cũng vì thế mà tăng cao. Dễ thấy được trong các lĩnh vực như: học tập, kinh doanh, giải trí, ăn uống,... đặc biệt là trong lĩnh vực du lịch. Cũng do ảnh hưởng bởi việc làm việc không ngừng nghỉ để phát triển nền kinh tế, vì thế mà nhu cầu trong việc nghỉ ngơi, đầu tư thời gian vào việc giải trí để “xả stress” của người dân cũng tăng cao.

Theo cách truyền thống, người dùng thường lựa chọn sản phẩm, các gói dịch vụ dựa trên sở thích, gợi ý của người thân hay những người dùng khác. Tuy nhiên với phương pháp truyền thống này có rất nhiều hạn chế khi phải lựa chọn quá nhiều, không phản ánh được sở thích của bản thân người dùng, mà nó giống một “phép thử” hơn, vì thế mà độ hài lòng trên từng sản phẩm cũng sẽ không thực tế.

Theo thống kê được đăng tải, Amazon năm 2018 có hơn 500 triệu sản phẩm được bán thuộc các lĩnh vực khác nhau. Netflix có hơn 30000 phim đang có sẵn trên trang này. Một con số rất lớn.

Cụ thể với ngành công nghiệp điện ảnh, theo thống kê của IMDb mỗi năm có hơn 10000 phim mới được phát hành trên toàn thế giới. Tại Việt Nam, chúng ta khoảng 50 phim ra rạp trong mỗi năm và khoảng hơn 100 phim tính trên toàn thế giới. Với các kênh truyền hình. Chúng ta có gần 200 kênh truyền hình được phát sóng, với tổng số lượng chương trình lên tới hơn một nghìn chương trình được phát thường xuyên. Trong khi đó, chúng ta có trung bình khoảng 2 tiếng cho việc giải trí, xem phim. Một con số quá ít với tổng số thời lượng của phim. Điều này gây rất nhiều khó khăn trong việc lựa chọn nội dung cho người dùng. Chúng ta không biết được bộ phim hay chương trình nào phù hợp với sở thích của mình. Với ngành du lịch nói riêng, năm 2018, nước ta đón 15,6 triệu lượt khách quốc tế, phục vụ trên 80 triệu lượt khách nội địa, tổng thu từ khách du lịch đạt hơn 620.000 tỷ đồng. Các địa phương là trung tâm du lịch lớn của cả nước như Hà Nội, thành phố Hồ Chí Minh, Quảng Ninh, Đà Nẵng... có tốc độ tăng trưởng mạnh. Số lượng khách sạn 4-5 tăng nhanh, công tác quảng bá, xúc tiến du lịch cũng được đẩy mạnh ở nhiều thị trường.

Với tiềm năng du lịch đa dạng và phong phú, Việt Nam là một nước có nhiều danh lam thắng cảnh, từ đó thu hút không ít khách du lịch cả trong và ngoài nước. Tuy nhiên, khách du lịch thường gặp phải rất nhiều khó khăn khi đi đến những thành phố lạ lẫm. Họ cần phải nhờ đến sự trợ giúp của hướng dẫn viên du lịch hoặc ít nhất là phải dựa

vào sách hướng dẫn hay bản đồ để có được những thông tin mà họ cần tìm. Tuy nhiên những sự trợ giúp này cũng có những hạn chế nhất định như: Khách du lịch sẽ có thể khó tìm thấy những thông tin chi tiết về những địa điểm du lịch cần tham quan để có sự chuẩn bị cần thiết. Hoặc là, những hướng dẫn viên du lịch thường hướng dẫn cho cả đoàn khách, họ đi theo những tour du lịch đã được định sẵn và chỉ được tham quan những điểm du lịch chính, trong khi những điểm du lịch thú vị lại không được tham quan, mặc dù cách địa điểm chính rất gần. Như vậy, làm thế nào để hỗ trợ khách khi du lịch có thể dễ dàng tìm thấy những địa điểm phù hợp với họ, phù hợp với những điều kiện ngữ cảnh xung quanh (như thời tiết, tâm trạng, bạn đồng hành,...).

Một số công cụ tìm kiếm với từ khóa, điển hình như Google, Bing, Cốc Cốc... hay các công cụ tìm kiếm trong các website du lịch giúp ta rất nhiều trong việc tìm kiếm các nội dung. Nhưng với lượng thông tin tăng lên nhanh chóng, các công cụ tìm kiếm theo từ khóa cũng đang dần tỏ ra không hoàn toàn đáp ứng được nhu cầu của chúng ta, khi các nội dung ngày một tăng dần đến sự trùng lặp về nội dung. Hơn nữa việc đôi khi người dùng không thể mô tả được nội dung, sở thích của mình thành từ khóa cụ thể để search. Do đó việc tìm kiếm phim phù hợp gặp rất nhiều khó khăn.

Vậy làm sao để người dùng có thể tiếp cận được đến những nội dung phù hợp với mình?

Để đáp ứng được nhu cầu người dùng có thể tìm được địa điểm du lịch nhanh chóng, phù hợp với sở thích, hoàn cảnh thì tư tưởng về hệ thống khuyến nghị được ra đời. Hệ thống đóng vai trò trung gian như một nhà tư vấn hỗ trợ người dùng đưa ra quyết định. Hệ thống này có khả năng thu thập dữ liệu người dùng, sau đó sử dụng dữ liệu đó để phân tích và giúp người dùng đưa ra quyết định dựa trên dữ liệu cá nhân của mỗi người. Kết hợp vào đó là các yếu tố ngữ cảnh có thể sẽ cho được kết quả phù hợp nhất.

Khóa luận đưa ra một giải pháp xây dựng hệ thống gợi ý dựa trên ngữ cảnh, sau đó ứng dụng giải pháp này cho bài toán gợi ý các điểm du lịch. Hệ thống được vận hành trên nền Web để hỗ trợ khách du lịch nhằm đem lại sự thoải mái và sự tiện dụng tối đa cho họ khi tham gia vào hệ thống. Để thực hiện, em đã nghiên cứu các phương pháp hiện có và đề xuất phương pháp xây dựng hệ thống bằng cách kết hợp phương pháp gợi ý dựa trên ngữ cảnh vào (contextual pre-filtering), tích hợp với kỹ thuật phân rã ma trận (matrix factorization) và xử lý ngữ cảnh đầu ra (contextual post-filtering), đồng thời tìm hiểu và đề xuất hướng khắc phục vấn đề người dùng mới trong hệ thống. Sau khi xây dựng hệ thống hoàn chỉnh, em thu thập ý kiến từ người dùng thực, từ đó đánh giá hiệu quả của hệ thống đã xây dựng.

## **Chương 2: Cơ sở lý luận**

Nội dung chính của chương gồm: giới thiệu tổng quan về đề tài, phạm vi của đề tài cùng các phương pháp tiếp cận vấn đề. Trong chương này trình bày thêm về bố cục của khóa luận và tổng quan về hệ thống gợi ý.

### **2.1 Giới thiệu**

#### **2.1.1. Đề tài**

Nghiên cứu hệ thống gợi ý dựa trên ngữ cảnh. Và ứng dụng hệ thống gợi ý vào các hệ thống online giúp nâng cao chất lượng trải nghiệm của người dùng. Cụ thể ở đây là hệ thống gợi ý địa điểm du lịch. Hệ thống gợi ý giúp giải quyết vấn đề làm cách nào để người dùng có thể chọn được địa điểm du lịch phù hợp với sở thích của mình tại một ngữ cảnh nhất định (gần với người dùng nhất,...). Như vậy việc đưa ra gợi ý phải dựa trên cơ sở cá nhân của mỗi người để lựa chọn ra những địa điểm thích hợp nhất trong hàng trăm nghìn địa điểm có trong hệ thống.

Yêu cầu đặt ra phải xây dựng hệ thống có khả năng gợi ý các địa điểm có khả năng phù hợp với sở thích, ngữ cảnh của người dùng.

Để làm rõ hơn các yêu cầu, chúng ta sẽ thu hẹp phạm vi nghiên cứu bằng cách trả lời các câu hỏi: Hệ thống gợi ý dựa trên phương pháp nào? Và tính chính xác của hệ thống gợi ý là bao nhiêu?

#### **2.1.2. Mục tiêu**

Mục tiêu của đề tài là nghiên cứu hệ thống gợi ý dựa trên ngữ cảnh: tầm quan trọng của hệ thống, các thức hoạt động, mô hình, phương pháp của hệ gợi ý.

Xây dựng và triển khai hệ thống gợi ý cùng với website có khả năng đưa ra những địa điểm du lịch cho người dùng tham khảo.

#### **2.1.3. Phạm vi**

Phạm vi nghiên cứu của đề tài bao gồm:

- Hệ khuyến nghị và ứng dụng triển khai trên nền tảng website để đưa ra những địa điểm du lịch phù hợp cho người dùng tham khảo.
- Đối tượng gợi ý: Các địa điểm du lịch trong cơ sở dữ liệu.
- Đối tượng được gợi ý: Người dùng có tài khoản (hoặc truy cập) trên hệ thống và có dữ liệu được lưu trên hệ thống.
- Nội dung gợi ý: Là tập các địa điểm được hệ thống cho là phù hợp nhất với người dùng.

#### **2.1.4. Phương pháp tiếp cận**

Trong khóa luận, hệ thống gợi ý sử dụng phương pháp lọc trước, bộ dữ liệu đánh giá địa điểm du lịch Tourpedia.

### **2.2. Tổng quan về hệ thống gợi ý**

Trong cuộc sống thường ngày, có rất nhiều trường hợp người ta phải đưa ra các lựa chọn dựa trên những ý kiến, lời khuyên của mọi người xung quanh, có thể là qua lời nói, sách báo, các bản đánh giá sản phẩm, v.v.. Nhưng trong kỷ nguyên internet, kỷ nguyên của thông tin, hàng triệu thông tin được đưa lên mạng mỗi ngày, điều này dẫn tới yêu cầu phải có các phương pháp tự động thu thập thông tin và đưa ra lời khuyên để hỗ trợ các phương pháp truyền thống trên. Hệ thống gợi ý là một giải pháp như vậy. Hệ thống này đưa ra gợi ý dựa trên những gì người dùng đã làm trong quá khứ, hoặc dựa trên tổng hợp ý kiến từ những người dùng khác. Hệ thống gợi ý đã trở thành một ứng dụng quan trọng và thu hút được sự quan tâm lớn của các nhà nghiên cứu cũng như các doanh nghiệp.

#### **2.2.1. Hệ thống gợi ý truyền thống**

Hệ thống gợi ý là một công cụ cung cấp một cách tự động và thông minh các gợi ý về sản phẩm, thực thể mà người dùng có thể quan tâm, hoặc các sản phẩm liên quan đến sản phẩm mà người dùng đang xem. Hệ thống gợi ý cũng được gọi là hệ thống hỗ trợ quyết định cho người dùng có thể mua sản phẩm, quan tâm đến một thực thể nào đó. Để đưa ra gợi ý cho người dùng, hệ thống sẽ dựa trên những đánh giá, những lần mua sắm trước, hoặc dựa trên câu hỏi của người dùng và kho tri thức.

Hệ thống gợi ý được áp dụng phổ biến trong rất nhiều các hệ thống như Youtube (gợi ý video người dùng có thể quan tâm), Amazon (gợi ý sản phẩm), Netflix (gợi ý phim) đóng vai trò quan trọng trong việc nâng cao trải nghiệm người dùng cũng như mang lại nhiều lợi nhuận hơn.

Mặc dù phổ biến nhưng hệ thống gợi ý vẫn đang được quan tâm để cải thiện các thuật toán nhằm tăng hiệu năng xử lý của hệ thống, và quan trọng hơn đó chính là tăng tính hiệu quả, độ chính xác trong gợi ý. Điển hình là rất nhiều contest được tổ chức trong đó có contest Netflix Prize do Netflix tổ chức với mục đích nâng cao tính chính xác của thuật toán.

#### **2.2.2. Hệ thống gợi ý trên ngữ cảnh**

Hệ tư vấn truyền thống vẫn chỉ quan tâm tới hai đối tượng của hệ tư vấn là người dùng (user) và sản phẩm (item) để đưa ra dự đoán một danh sách ngắn các sản phẩm mà người dùng có thể quan tâm. Tuy nhiên trên thực tế, sở thích của người dùng lại

không cố định. Ví dụ một người lúc buồn thích nghe những bản nhạc R&B, bolero, lúc vui lại thích nghe indie, rap battle. Lúc trời mưa thích đi chơi ở siêu thị, lúc khô sẽ đi chơi ở công viên, đi phượt, leo núi. Có thể nói sở thích của người dùng bị tác động nhiều bởi những yếu tố ngữ cảnh bên ngoài. Như vậy, nếu ta sử dụng dữ liệu của ngữ cảnh để phân tích sẽ cho kết quả tốt hơn hệ tư vấn truyền thống.

### 2.2.3. Bài toán gợi ý trên ngữ cảnh

Bài toán gợi ý được mô tả như sau

Gọi tập (U) là tập tất cả các người dùng, (I) là tập các sản phẩm có thể được gợi ý. Kích thước của các tập đều rất lớn, có thể từ vài trăm nghìn (hệ thống phim, sách, nhạc, v.v..) cho đến vài triệu sản phẩm (hệ thống thương mại điện tử, v.v.).

Tập các ngữ cảnh (C) thể hiện các ngữ cảnh ảnh hưởng đến đánh giá của người dùng. Hàm  $r(u, i, c)$  là dự đoán đánh giá của người dùng  $u$  với sản phẩm  $i$  trong điểm ngữ cảnh  $c$ . Nhiệm vụ của hệ thống gợi ý là tìm ra sản phẩm  $i' \in (I)$  cho người dùng  $u' \in (U)$  tại  $c$  sao cho hàm  $r(u, i, c)$  đạt giá trị lớn nhất.

Trong hệ thống gợi ý, đánh giá của người dùng với mỗi một sản phẩm được đo bằng điểm, ví dụ người dùng A đánh giá bộ phim “Avengers: End game” với điểm số là 4/5 điểm. Điểm số này thể hiện mức độ hài lòng của người dùng đó với bộ phim. Thông qua đó phân nào thể hiện được sở thích cũng người dùng.

Mỗi người dùng trong không gian (U) được biểu diễn bằng một hồ sơ người dùng. Trong hồ sơ này định danh với mỗi người dùng là ID của người dùng (userid). Tương tự với mỗi địa điểm giải trí - du lịch trong không gian (I) được xác định bởi một mã số sản phẩm (itemid) và các đặc trưng của địa điểm đó. Ví dụ trong hệ thống gợi ý địa điểm du lịch, các đặc trưng có thể là: tên địa điểm, vị trí, thể loại, v.v..

Trong hệ thống gợi ý địa điểm du lịch, hàm  $r$  không được xác định trên toàn không gian  $(U) \times (I) \times (C)$  mà chỉ xác định trong một miền nhỏ của không gian này (mỗi người dùng đánh giá một phần của tập (I)). Vì vậy hàm  $r$  phải dự đoán phần còn lại chưa được đánh giá của người dùng. Độ tương tự được tính toán dựa trên tập các sản phẩm đã được đánh giá trước đó. Từ những thông tin đó, hệ thống gợi ý phải dự đoán điểm cho các địa điểm chưa được người dùng đánh giá và đưa ra những gợi ý phù hợp nhất.

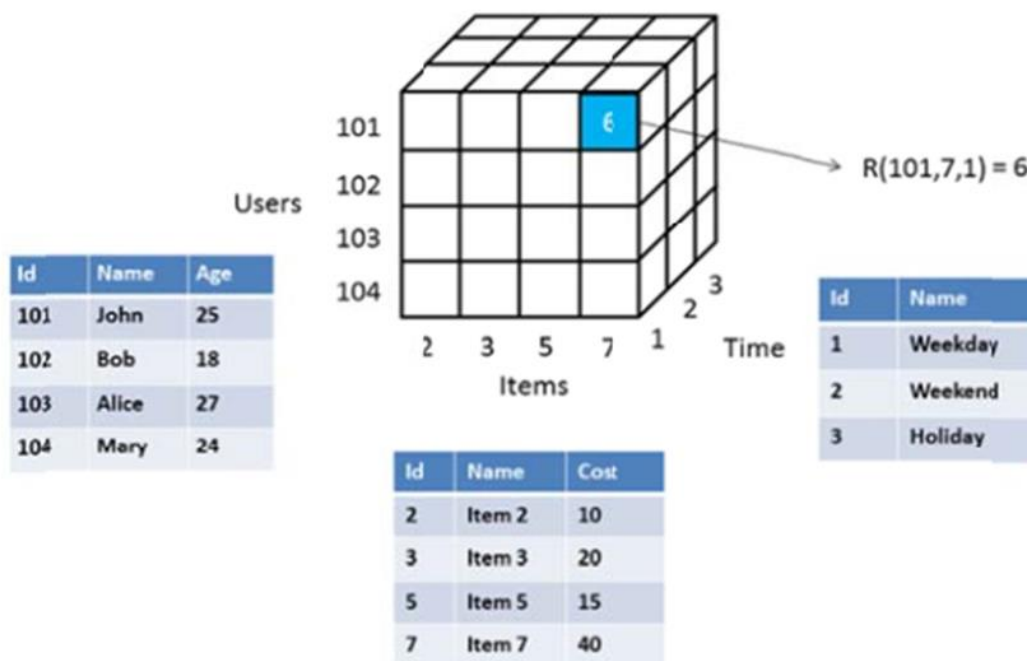
Userid	ItemId	Điểm	Thời gian	Thời tiết	Người đồng hành
1	2	1	Cuối tuần	Trời nắng	Một mình



1	5	1	Cuối tuần	Trời âm u	Gia đình
1	3	1	Lễ - tết	Trời trong xanh	Một mình
2	2	2	Ngày trong tuần	Trời nắng	Bạn bè
2	4	3	Lễ - tết	Trời trong xanh	Bạn bè
3	6	3	Cuối tuần	Trời nắng	Bạn bè
4	1	2	Lễ - tết	Trời mưa	Một mình

Bảng 2.1 Bảng dữ liệu đánh giá chứa thông tin ngữ cảnh

Không gian gợi ý ba chiều có thể được mô tả trong khối lập phương (Hình 2.1). Ô tô đậm cho biết chỉ số  $R(101,7,1) = 6$  có ý nghĩa là người dùng có mã số 101 đánh giá sản phẩm có mã số 7 trong điều kiện thời gian có mã số 1 với giá trị đánh giá là 6. Trong khối lập phương này, không phải ô nào cũng có giá trị, những ô không có giá trị là do người dùng chưa đánh giá. Mục tiêu của hệ thống gợi ý là dự đoán giá trị tại những ô còn thiếu đó, từ đó đưa ra lời gợi ý đến với người dùng.



Hình 2.1 Minh họa cho không gian gợi ý ba chiều

Các phương pháp tiếp cận trong hệ thống gợi ý đa chiều:

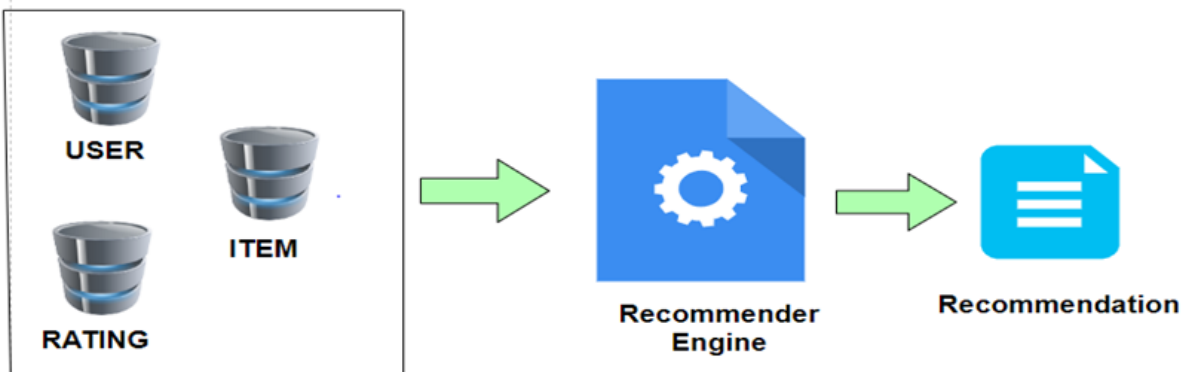
- Xử lý ngữ cảnh đầu vào (Contextual pre-filtering): Trong tiếp cận này, thông tin về ngữ cảnh được sử dụng để lựa chọn hoặc xây dựng các thiết

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 được 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.

#### 2.2.4. Cách thức hoạt động của hệ thống gợi ý.

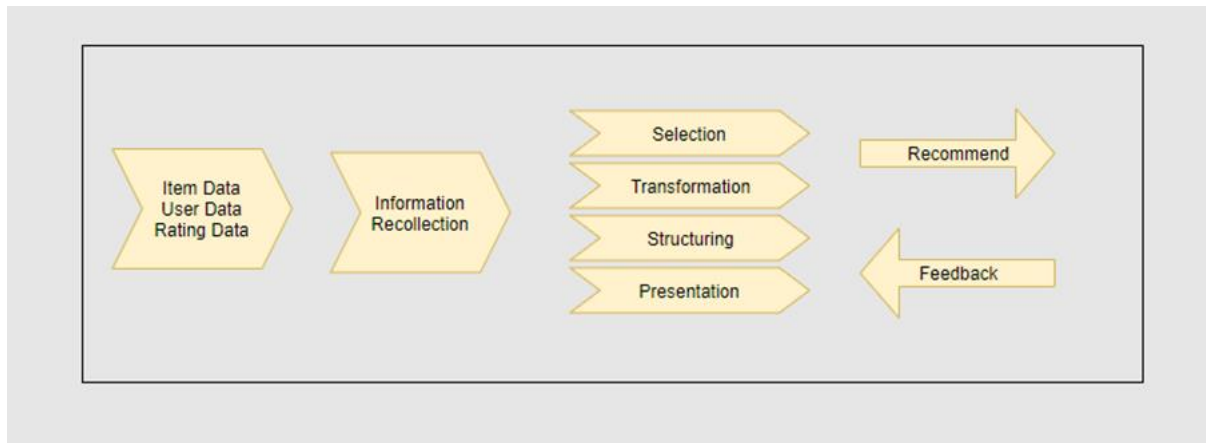
Các thành phần của hệ thống gợi ý được mô tả như hình bên dưới:



Hình 2.2. Các thành phần chính của hệ thống gợi ý

- Thông tin người dùng (User Data): thông tin về người dùng như là danh tính, email, username, password, v.v..
- Thông tin về sản phẩm (Item Data): các thông tin về sản phẩm là kết quả gợi ý cho người dùng. Bao gồm các đặc tính của sản phẩm. Có thể là các địa điểm giải trí – du lịch, bản nhạc, bài báo, sách, tạp chí, v.v..
- Thông tin về đánh giá (Rating Data): là các thông tin đầu vào của việc gợi ý. Dữ liệu là số điểm đánh giá của người dùng về địa điểm giải trí – du lịch, ngành hàng trong hệ thống.

- Bộ máy gợi ý (Recommender Engine): là chương trình thực hiện phân tích, tính toán dữ liệu dựa trên các thông tin có sẵn. Dựa vào các thuật toán nhất định để đưa được ra các kết quả tốt nhất với người dùng.
- Kết quả gợi ý (Recommendation): là kết quả của việc tính toán của bộ máy gợi ý. Kết quả đó chính là thông tin của các sản phẩm trong không gian sản phẩm được cho là phù hợp nhất với người dùng.
  - Quá trình hoạt động của hệ thống được mô tả chi tiết dưới hình như sau:



*Hình 2.3 Quá trình hoạt động chi tiết của hệ thống*

Quá trình hoạt động của hệ thống bao gồm các bước như sau:

**Tập hợp thông tin (Information Recollection):** Bao gồm các thông tin cá nhân người sử dụng, lịch sử sử dụng các sản phẩm, các địa điểm đã du lịch (tự khai), lịch sử các giao dịch, thông tin về sự ưu tiên của người dùng (user preference), và thông tin về sản phẩm như đặc tính sản phẩm. Bước này đóng một vai trò quan trọng vì nó là cơ sở cho toàn bộ quá trình hệ gợi ý thực hiện dự đoán và đưa ra gợi ý cho người dùng. Nếu thông tin thu thập không đầy đủ, có sai sót hay mâu thuẫn hệ thống gợi ý sẽ dự đoán sai hoặc không thể thực hiện chức năng dự đoán.

**Lựa chọn thông tin (Selection):** Bao gồm việc xác định những thông tin liên quan trực tiếp đến quá trình xử lý, dự đoán. Cách lựa chọn tập thông tin phụ thuộc vào phương pháp tiếp cận của hệ thống. Từ tập thông tin lựa chọn giúp chúng ta xác định được độ tương tự giữa hai sản phẩm bất kì hoặc độ tương tự giữa hai người dùng.

**Chuyển đổi thông tin (Transformation):** mục tiêu chính của bước chuyển đổi là thực hiện các biến đổi thông tin, xây dựng các mô hình, các hàm từ dữ liệu đã được xử lý trong hai bước đầu, biểu diễn thông tin dưới dạng quy định của hệ gợi ý để có thể thực hiện các dự đoán.

Cấu trúc thông tin (Structuring): Cấu trúc thông tin có liên quan đến việc tổ chức thông tin khi người dùng xem qua các thông tin gợi ý đưa ra. Bước này bao gồm các hoạt động như phân nhóm các sản phẩm, phân loại, xếp hạng các sản phẩm...

Trình bày thông tin (Presentation): bước cuối cùng trong quá trình gợi ý là đưa ra các thông tin gợi ý cho người dùng theo các tiêu chuẩn: sắp xếp, bố trí, định dạng tài liệu, màu sắc, phong chữ... Đây là bước cuối cùng trong quá trình gợi ý, đưa ra kết quả cho người dùng cụ thể.

Thông tin phản hồi: Thu nhận các thông tin phản hồi là một bước thêm vào, tùy theo mỗi hệ thống gợi ý. Bước này giúp ích rất nhiều trong việc cải thiện hoạt động, nâng cao kết quả của hệ gợi ý. Phản hồi của người dùng được hệ thống ghi nhận với hai dạng tiềm ẩn (implicit), hoặc rõ ràng (explicit):

- Phản hồi rõ ràng (explicit feedback): người sử dụng cung cấp các thông tin thể hiện mức độ ưu tiên với các sản phẩm liên quan như đánh giá, cho điểm sản phẩm.
- Phản hồi tiềm ẩn (implicit feedback): có được bằng cách thu thập, phân tích các hành vi người dùng như: lịch sử duyệt các sản phẩm, số lần viếng thăm, thời gian lưu lại...

#### **2.2.5. Tính ứng dụng của hệ thống gợi ý**

Hệ thống gợi ý là công cụ phần mềm có khả năng đưa ra các gợi ý cho người dùng về các sản phẩm, dịch vụ, phim ảnh, ca nhạc mà có thể họ thích dựa trên thói quen và sở thích của mỗi người. Dưới đây là một số lợi của hệ thống gợi ý

Tăng số lượng các mặt hàng bán ra cho hệ thống thương mại điện tử. Đây là chức năng quan trọng nhất của hệ thống gợi ý. Thay vì người dùng chỉ mua một sản phẩm mà họ cần, họ được gợi ý mua những sản phẩm “có thể họ quan tâm” mà bản thân họ không nhận ra. Hệ thống gợi ý tìm ra những “mối quan tâm ẩn” qua các tương tác của người dùng. Bằng cách đó, hệ thống gợi ý làm gia tăng nhu cầu của người dùng, mở rộng thông tin và gia tăng số lượng mặt hàng bán ra.

Tăng độ tin cậy của người dùng: Khi hệ thống gợi ý cho người dùng những lựa chọn và họ hài lòng về những gợi ý đó thì lòng tin của họ đối với hệ thống sẽ được nâng lên. Hệ thống gợi ý hoạt động dựa trên những xếp hạng thật từ chính bản thân người dùng trong quá khứ. Do đó, khi người dùng càng tin cậy vào hệ thống, đưa ra những đánh giá trung thực cho các sản phẩm, hệ thống sẽ mang lại cho người dùng nhiều gợi ý chính xác hơn, phù hợp với nhu cầu, sở thích của họ.

Tăng sự hài lòng của người dùng: Hệ thống gợi ý được xây dựng để tìm hiểu nhu cầu, sở thích của người dùng, gợi ý cho họ những thứ họ có thể cần. Chính vì vậy hệ thống gợi ý tăng sự hài lòng của người dùng đối với hệ thống.

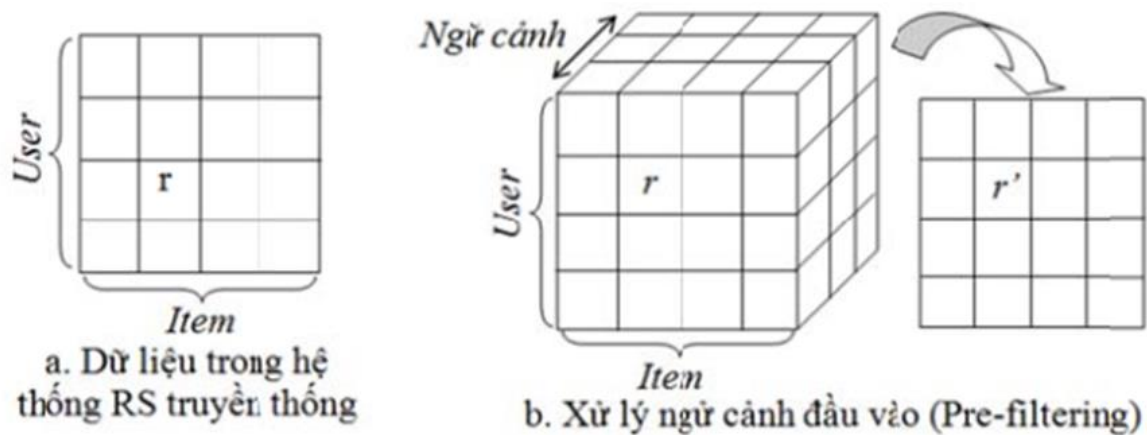
### **Kết luận**

Như đã trình bày ở trên hệ thống gợi ý mang lại rất nhiều lợi ích cho các hệ thống thương mại, dịch vụ,... Để có thể xây dựng được một hệ thống gợi ý chúng ta phải giải quyết được bài toán đã đặt ra. Chương 3 sẽ trình bày các thuật toán để giải quyết bài toán về hệ thống gợi ý.

## Chương 3: Thực trạng

Nội dung chính của chương này là trình bày một số các phương pháp để giải quyết bài toán gợi ý dựa trên ngữ cảnh. Trình bày về cách thức dự đoán của thuật toán phân rã ma trận.

### 3.1. Phương pháp lọc ngữ cảnh trước



Hình 3.1 Phương pháp lọc ngữ cảnh trước

Khi phương pháp lọc trước khi gợi ý, hệ thống sẽ lấy ngữ cảnh hiện tại của người dùng để so sánh với từng đánh giá đã có trong ngữ cảnh như thế, từ đó loại bỏ những đánh giá ở ngữ cảnh khác. Ví dụ như người dùng đăng nhập vào cuối tuần và sắp xem phim với bạn bè, hệ thống sẽ loại bỏ tất những đánh giá với điều kiện ngữ cảnh người dùng đăng nhập vào thời điểm trong tuần. Sau đó bài toán sẽ trở thành bài toán gợi ý truyền thống với hai đối tượng là người dùng và sản phẩm.

Trong phương pháp này, dữ liệu ban đầu được mô tả bằng ma trận nhiều chiều. Số lượng chiều từ chiều thứ 3 phụ thuộc vào số ngữ cảnh mà ta đang xét. Trong phạm vi của khóa luận, ta nghiên cứu trên 3 chiều ngữ cảnh là Time, Weather, Companion.

Như vậy để giải quyết bài toán gợi ý này, ta cần chuyển đổi ma trận nhiều chiều User-Item-Context thành ma trận ma trận hai chiều User-Item để thực hiện tính toán độ tương tự

Bảng dưới đây mô tả một mảnh nhỏ dữ liệu đánh giá của người dùng trong hệ thống gợi ý có kết hợp yếu tố ngữ cảnh.

<b>1</b>	<b>Userid</b>	<b>Itemid</b>	<b>Rating</b>	<b>Time</b>	<b>Weather</b>	<b>Companion</b>
0	u1	i1	2	Weekday	Sunny	Family
1	u1	i1	4	Holiday	Rainy	Alone
2	u1	i1	5	Weekday	Cloudy	Partner
3	u1	i1	4	Weekday	Rainy	
4	u1	i1	1	Holiday	Cloudy	Family
5	u1	i1	2	Weekday	Sunny	Alone
6	u1	i1	3	Weekend	Nice	Partner
7	u1	i1	2	Weekday	Sunny	
8	u1	i1	1	Holiday	Sunny	Family
9	u1	i1	1	Weekday		Alone
10	u1	i1	2	Weekend		Partner
11	u1	i1	3	Weekend		
12	u1	i1	4	Weekend	Sunny	Family
13	u1	i1	?	Holiday	Sunny	Alone
14	u1	i1	2	Weekend	Cloudy	Partner
15	u1	i1	3	Holiday	Rainy	
16	u1	i1	?	Weekend	Sunny	Family
17	u1	i1	2	Weekend	Rainy	Alone
18	u1	i1	3	Weekday	Sunny	Partner
19	u1	i1	?	Weekend	Cloudy	
20	u1	i1	2	Weekday		Family
21	u1	i1	2	Holiday		Alone

22	u1	i1	?	Weekday		Partner
23	u1	i1	?	Weekend		
24	u1	i1	?			Family
25	u1	i1	3			Alone
26	u1	i1	3			Partner
27	u1	i1	4			

*Bảng 3.1 Bảng đánh giá người dùng với địa điểm du lịch trong các điều kiện ngữ cảnh*

Đối với hệ thống lọc ngữ cảnh trước, khi người dùng đăng nhập vào ngày lễ, trời nắng và đi với gia đình. Khi đó dữ liệu sẽ được lọc dựa trên ba điều kiện của ngữ cảnh là Time = Holiday, Weather = Sunny, Companion = Family. Như trong mảnh dữ liệu trên hệ thống gợi ý sẽ lọc ra được bản ghi của người dùng u1 với item i1 như sau.

Index	UserId	ItemId	Rating	Time	Weather	Companion
8	u1	i1	1	Holiday	Sunny	Family

*Bảng 3.2 Bảng dữ liệu của một người dùng sau khi sử dụng phương pháp lọc trước*

Như vậy sau khi sử dụng lọc ngữ cảnh trước, tất cả các ngữ cảnh không thuộc ngữ cảnh hiện tại sẽ không được đưa vào bộ data đang xét làm đầu vào dữ liệu để tính toán độ tương tự. Như vậy từ ma trận đánh giá nhiều chiều chúng ta đã đưa được về ma trận đánh giá hai chiều để thực hiện tính toán độ tương tự.

userid	itemid	Rating	Time	Weather	Companion
1067	tt0356910	1	Holiday	Sunny	Family
1067	tt3793764	1	Holiday	Sunny	Family
1067	tt0111161	1	Holiday	Sunny	Family



1067	tt0120338	1	Holiday	Sunny	Family
1067	tt1099212	1	Holiday	Sunny	Family
1018	tt0110357	4	Holiday	Sunny	Family
1018	tt0111161	5	Holiday	Sunny	Family
1018	tt0169547	5	Holiday	Sunny	Family
1018	tt0367594	5	Holiday	Sunny	Family
1018	tt0133093	5	Holiday	Sunny	Family
1009	tt0454876	4	Holiday	Sunny	Family
1009	tt0181689	4	Holiday	Sunny	Family
1009	tt1453405	5	Holiday	Sunny	Family
1009	tt0993846	4	Holiday	Sunny	Family
1009	tt0211915	5	Holiday	Sunny	Family
1057	tt1637706	2	Holiday	Sunny	Family
1057	tt0378194	5	Holiday	Sunny	Family
1057	tt0376541	1	Holiday	Sunny	Family
1057	tt0088763	4	Holiday	Sunny	Family
1057	tt0133093	3	Holiday	Sunny	Family
1074	tt0356910	4	Holiday	Sunny	Family
1074	tt1657301	3	Holiday	Sunny	Family
1074	tt2251217	4	Holiday	Sunny	Family
1074	tt0114369	1	Holiday	Sunny	Family

1074	tt0993846	1	Holiday	Sunny	Family
------	-----------	---	---------	-------	--------

*Bảng 3.3 Dữ liệu đánh giá của người dùng sau khi lọc ngữ cảnh*

Sau khi lọc được toàn bộ các dữ liệu theo chiều ngữ cảnh xác định, ta có thể dễ dàng chuyển đổi dữ liệu thành ma trận hai chiều User-Item thể hiện đánh giá của người dùng như trong phương pháp gợi ý truyền thống.

Bảng dưới đây thể hiện sự đánh giá của các User với các Item sau khi thực hiện lọc ngữ cảnh Holiday – Sunny – Family.

	tt0356910	tt3793764	tt0111161	tt0120338	tt1099212	tt0110357	tt0169547
1067	1	1	1	1	1	0	0
1018	0	0	5	0	0	4	5
1009	0	0	0	0	0	0	0
1057	0	0	0	0	0	0	0
1074	4	0	0	0	0	0	0

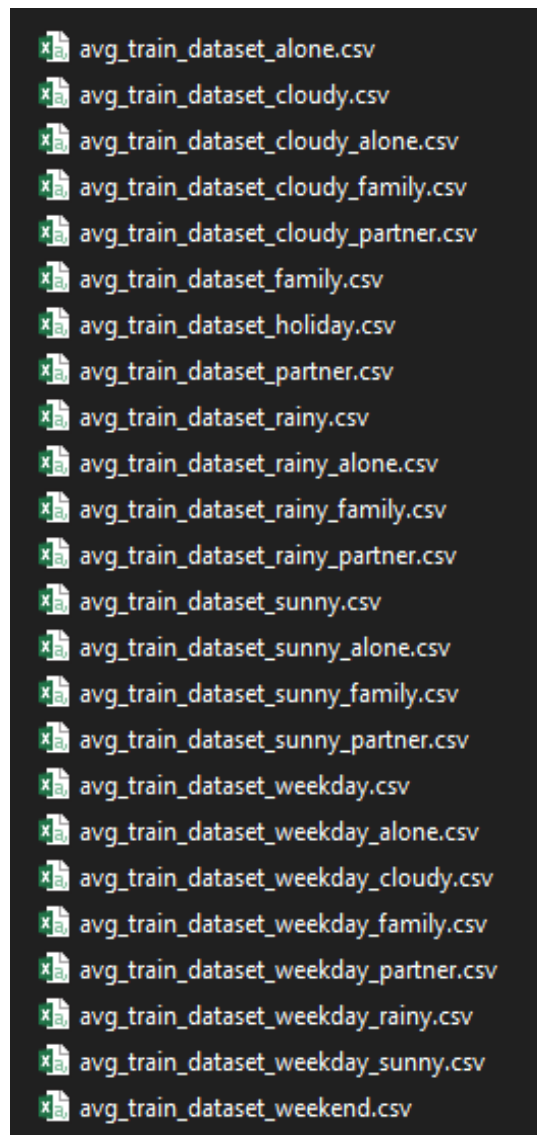
*Bảng 3.4 Bảng dữ liệu User-Item sau khi lọc ngữ cảnh*

Như vậy, với mỗi điểm ngữ cảnh, ta sẽ có một tập dữ liệu riêng cho từng điểm ngữ cảnh đó. Rõ ràng, chúng ta không thể biết trước được ngữ cảnh mà người dùng yêu cầu gợi ý. Nếu như mỗi khi người dùng gửi request lấy gợi ý ta thực hiện lọc ngữ cảnh ngay lúc đó, chuẩn hóa, tính toán lại dữ liệu thì sẽ gây ra rất nhiều vấn đề cho hệ thống. Vấn đề đầu tiên ta có thể nhìn thấy đó là thời gian phản hồi của hệ thống sẽ cực kì thấp bởi vì chi phí để chuẩn hóa dữ liệu cũng như là tính toán độ tương tự là rất lớn. Vấn đề thứ hai là không thể đáp ứng cho nhiều người dùng nếu như ta thực hiện tính toán ngay khi người dùng yêu cầu. Do vậy để giải quyết hai vấn đề trên, ta phải có phương pháp để chuẩn bị dữ liệu đã được tính toán cho người dùng.

Trong nghiên cứu này, khóa luận đề xuất một phương pháp chuẩn bị chia tách dữ liệu thành các mảnh nhỏ, mỗi mảnh ứng với mỗi điểm ngữ cảnh khác nhau. Giả sử với trường hợp người dùng yêu cầu gợi ý tại điểm ngữ cảnh là Holiday – Weather – F. Ta cần chuẩn bị sẵn tập dữ liệu đã được lọc theo điểm ngữ cảnh này và thực hiện huấn luyện. Khi đó bộ dữ liệu này sẽ là độc lập với các dữ liệu ở ngữ cảnh khác và đã được tính toán độ tương tự của các user hay item trong hệ thống tùy thuộc vào cách xây dựng hệ thống.

Tuy nhiên nhược điểm của giải pháp này là sẽ tạo ra nhiều mảnh dữ liệu. Số lượng mảnh dữ liệu sẽ phụ thuộc vào các điều kiện ngữ cảnh mà ta đang xét đến trong hệ thống. Khi đó, số lượng mảnh dữ liệu được sinh ra sẽ là tổ hợp của các điều kiện ngữ cảnh. Nhưng với điều kiện thực tế, số lượng điều kiện ngữ cảnh mà các hệ thống xét đến thường không quá lớn. Do đó phương pháp này mang lại hiệu quả khá cao về việc cải thiện hiệu năng của hệ thống.

Dưới đây là ví dụ về các mảnh dữ liệu trong phạm vi nghiên cứu của khóa luận.



Hình 3.2 Các mảnh dữ liệu ngữ cảnh

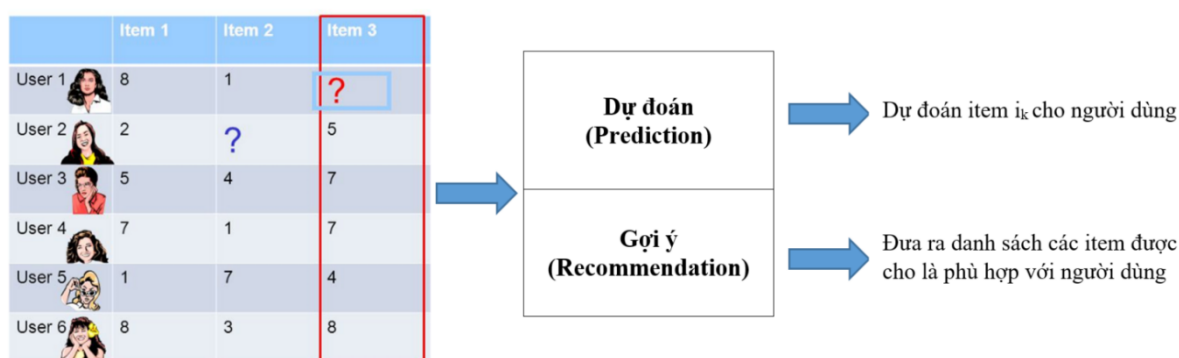
## Giải thuật lọc cộng tác

Lọc cộng tác là một giải thuật mạnh và được cho là đem lại hiệu quả cao cho nhiều hệ thống gợi ý. Về cơ bản, hệ lọc cộng tác sẽ tự động hóa việc gợi ý dưới hình thức tính toán dựa vào sở thích của những người dùng trước trong hệ thống nhằm dự đoán mối quan tâm của người dùng đó với sản phẩm mới, hoặc gợi ý cho họ về những sản phẩm này.

Trong hệ lọc cộng tác cụ thể, chúng ta đưa ra tập  $U$  người dùng và một tập sản phẩm. Với mỗi người dùng  $u_i$  có thể đưa ra đánh giá về một tập sản phẩm  $I_{ui}$ . Hệ lọc cộng tác đưa ra những sản phẩm được cho là phù hợp với nhất với người dùng dựa hai yếu tố sau:

- Dự đoán (Prediction) là các giá trị bằng số, biểu thị giá trị đã được dự đoán cho item mà người dùng gợi ý. Các giá trị được dự đoán phải nằm trong khoảng cho phép (tùy theo quy ước của hệ thống đánh giá, trong khóa luận này là từ 1 – 5).
- Gợi ý (Recommendation) là danh sách các items, mà người dùng có thể quan tâm đến. Danh sách này phải nằm trong những item chưa được người dùng đánh giá.

Dưới đây là hình thể hiện tiến trình của hệ lọc cộng tác



Hình 3.3 Tiến trình lọc cộng tác

Thuật toán CF biểu diễn toàn bộ dữ liệu của ma trận  $U \times I$  (user-item) như là một ma trận trọng số đánh giá  $A$ . Mỗi thực thể  $a$  trong  $A$  là một điểm thể hiện mối quan tâm (hay chính là đánh giá) của một người dùng về một sản phẩm. Các đánh giá này được đặt trong một khoảng quy định trước và cũng có thể bằng 0 nếu người dùng đó chưa từng biết đến (chưa đánh giá) sản phẩm đó.

### ❖ Kỹ thuật phân rã ma trận

Mỗi item được mô tả bằng vector  $\mathbf{x}$  được gọi là item profile. Trong phương pháp này, ta cần tìm một vector hệ số  $\mathbf{w}$  tương ứng với mỗi user sao cho rating đã biết mà user đó cho item xấp xỉ với:

$$y \approx \mathbf{xw}$$

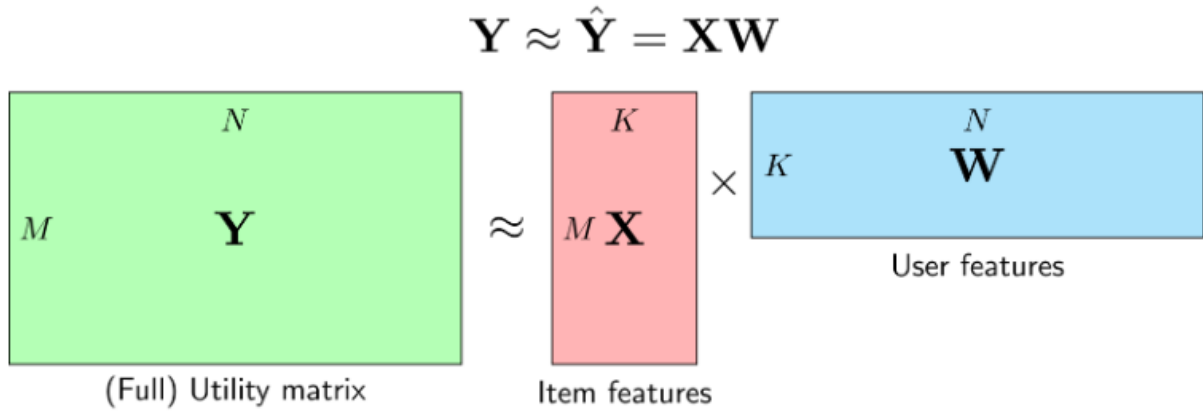
Với cách làm trên, *Utility Matrix*  $\mathbf{Y}$  (ma trận đánh giá), giả sử đã được điền hết, sẽ xấp xỉ với:

$$\mathbf{Y} \approx \begin{bmatrix} \mathbf{x}_1 \mathbf{w}_1 & \mathbf{x}_1 \mathbf{w}_2 & \dots & \mathbf{x}_1 \mathbf{w}_N \\ \mathbf{x}_2 \mathbf{w}_1 & \mathbf{x}_2 \mathbf{w}_2 & \dots & \mathbf{x}_2 \mathbf{w}_N \\ \dots & \dots & \ddots & \dots \\ \mathbf{x}_M \mathbf{w}_1 & \mathbf{x}_M \mathbf{w}_2 & \dots & \mathbf{x}_M \mathbf{w}_N \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \dots \\ \mathbf{x}_M \end{bmatrix} [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \dots \quad \mathbf{w}_N] = \mathbf{XW}$$

với  $M, N$  lần lượt là số items và số users

Ở ma trận trên,  $\mathbf{X}$  được xây dựng dựa trên thông tin mô tả của *item* và quá trình xây dựng này độc lập với quá trình đi tìm hệ số phù hợp cho mỗi *user*. Như vậy, việc xây dựng *item profile* đóng vai trò rất quan trọng và có ảnh hưởng trực tiếp lên hiệu năng của mô hình. Thêm nữa, việc xây dựng từng mô hình riêng lẻ cho mỗi *user* dẫn đến kết quả chưa thực sự tốt vì không khai thác được đặc điểm của những *users* gần giống nhau.

Bây giờ, giả sử rằng ta không cần xây dựng từ trước các *item profile*  $\mathbf{x}$  mà vector đặc trưng cho mỗi *item* này có thể được huấn luyện đồng thời với mô hình của mỗi *user* (ở đây là 1 vector hệ số). Điều này nghĩa là, biến số trong bài toán tối ưu là cả  $\mathbf{X}$  và  $\mathbf{W}$ ; trong đó,  $\mathbf{X}$  là ma trận của toàn bộ *item profiles*, mỗi **hàng** tương ứng với 1 *item*,  $\mathbf{W}$  là ma trận của toàn bộ *user models*, mỗi **cột** tương ứng với 1 *user*. Như vậy, *Utility Matrix*  $\mathbf{Y} \in \mathbb{R}^{M \times N}$  bằng tích của hai ma trận  $\mathbf{X} \in \mathbb{R}^{M \times K}$  và  $\mathbf{W} \in \mathbb{R}^{K \times N}$  với  $K \ll M, N$



Hình 3.4 Utility matrix  $Y$  được phân tích thành tích hai ma trận low rank  $X$  và  $W$

Ý tưởng của thuật toán dựa trên sự tồn tại của các *latent features* (tính chất ẩn) – mô tả sự liên quan giữa các items và users. Ví dụ với hệ thống gợi ý các địa điểm du lịch, tính chất ẩn có thể là các loại địa điểm như: khu vui chơi giải trí, trung tâm thương mại, di tích lịch sử,... Mỗi item sẽ mang tính chất ẩn ở một mức độ nào đó tương ứng với các hệ số trong vector  $\mathbf{x}$  của nó, hệ số càng cao tương ứng với việc mang tính chất đó càng cao. Xu hướng thích tính chất ẩn đó ở mỗi user cũng được mô tả bởi các hệ số trong vector  $\mathbf{w}$ . Hệ số càng cao nếu các thành phần tương ứng của  $\mathbf{x}$  với  $\mathbf{w}$  đều cao. Dựa vào đó, ta sẽ biết được user thích item nào để gợi ý.

Kỹ thuật phân rã ma trận (*Matrix Factorization*) được xếp vào dạng Collaborative Filtering bởi vì hàm mất mát mà chúng ta sắp xây dựng cần tìm nghiệm để tối ưu, muốn vậy, ta phải lần lượt đi tìm  $\mathbf{X}$  và  $\mathbf{W}$  khi thành phần còn lại được cố định. Như vậy, mỗi hàng của  $\mathbf{X}$  sẽ phụ thuộc vào toàn bộ các cột của  $\mathbf{W}$ . Ngược lại, mỗi cột của  $\mathbf{W}$  lại phụ thuộc vào toàn bộ các hàng của  $\mathbf{X}$ . Như vậy, có những mối quan hệ ràng buộc *chằng chịt* giữa các thành phần của hai ma trận trên. Tức chúng ta cần sử dụng thông tin của tất cả để suy ra tất cả. Vậy nên phương pháp này cũng được xếp vào Collaborative Filtering.

Với kỹ thuật phân rã ma trận, việc learning có thể phức tạp hơn Neighborhood-based Collaborative Filtering, nhưng việc inference đơn giản hơn vì ta chỉ cần lấy tích của hai vector  $\mathbf{xw}$ , mỗi vector có độ dài  $\mathbf{K}$  là một số nhỏ hơn nhiều so với  $\mathbf{M}$ ,  $\mathbf{N}$ . Vậy nên quá trình *inference* không yêu cầu khả năng tính toán cao. Việc này khiến nó phù hợp với các mô hình có tập dữ liệu lớn.

Song song với đó, việc lưu trữ hai ma trận  $\mathbf{X}$  và  $\mathbf{W}$  yêu cầu lượng bộ nhớ nhỏ hơn. Với MF, bộ nhớ cần để chứa là  $\mathbf{K(M+N)}$  còn với NBCF là  $\mathbf{M^2}$  hoặc  $\mathbf{N^2}$

### ❖ Xây dựng hàm mất mát

Vì  $X$ ,  $M$  đều là một dải giá trị, nên hàm mất mát có thể tính được thông qua mô hình tuyến tính Linear Regression.

Giả sử rằng số *users* là  $N$ , số *items* là  $M$ , UT được mô tả bởi ma trận  $Y$ . Thành phần ở hàng thứ  $m$ , cột thứ  $n$  của  $Y$  là mức độ quan tâm (giá trị đã rate) của *user* thứ  $n$  lên sản phẩm thứ  $m$  mà hệ thống thu thập được. Ma trận  $Y$  bị khuyết rất nhiều thành phần tương ứng với các giá trị mà hệ thống cần dự đoán. Thêm nữa, gọi  $R$  là ma trận rated or not thể hiện việc một user đã rated một item hay chưa. Cụ thể,  $r_{ij}$  bằng 1 nếu *item* thứ  $i$  đã được rated bởi *user* thứ  $j$ , bằng 0 trong trường hợp ngược lại.

Giả sử rằng ta có thể tìm được một mô hình cho mỗi user, minh họa bởi vector cột hệ số  $w_i$  sao cho mức độ quan tâm của một user tới một item có thể tính được bằng một hàm tuyến tính:

$$y = \mathbf{x}_m \mathbf{w}_n$$

(Với item tương tự vector  $x$ )

Xét một user (item) thứ  $n$  ( $m$ ) bất kỳ, nếu ta coi training set là một tập hợp các thành phần đã được điền của  $y_{n(m)}$ , ta có thể xây dựng hàm mất mát tương tự như Ridge Regression như sau:

$$L(X, W) = \frac{1}{2s} \sum_{n=1}^N \sum_{m:r_{mn}=1} (y_{mn} - \mathbf{x}_m \mathbf{w}_n)^2 + \frac{\lambda}{2} (\|\mathbf{X}\|_F^2 + \|\mathbf{W}\|_F^2) \quad (1)$$

Trong đó  $\|\mathbf{X}\|_F^2$  là *Frobenius norm*, tức căn bậc hai của ma trận (giống với norm 2 trong vector),  $s$  là toàn bộ số *ratings* đã có. Thành phần thứ nhất là trung bình sai số của mô hình. Thành phần thứ hai trong hàm mất mát phía trên là  $l_2$  regularization, giúp tránh *overfitting*.

Giá trị ratings thường là các giá trị đã được chuẩn hóa, bằng cách trừ mỗi hàng của Utility Matrix đi trung bình cộng của các giá trị đã biết của hàng đó (item-based) hoặc trừ mỗi cột đi trung bình cộng của các giá trị đã biết trong cột đó (user-based). Trong một số trường hợp nhất định, ta không cần chuẩn hóa ma trận này, nhưng kèm theo đó phải có thêm các kỹ thuật khác để giải quyết vấn đề *thiên lệch* trong khi *rating*.

### ❖ Tối ưu hàm mất mát

Việc tối ưu đồng thời  $X$ ,  $W$  là tương đối phức tạp, thay vào đó, phương pháp được sử dụng là lần lượt tối ưu một ma trận trong khi cố định ma trận kia, tới khi hội tụ.

Khi cố định  $X$ , việc tối ưu  $W$  được đưa về tối ưu hàm:

$$L(W) = \frac{1}{2s} \sum_{n=1}^N \sum_{m:r_{mn}=1} (\mathbf{y}_{mn} - \mathbf{x}_m \mathbf{w}_n)^2 + \frac{\lambda}{2} \|\mathbf{W}\|_F^2 \quad (2)$$

Khi cố định  $\mathbf{W}$ , việc tối ưu  $\mathbf{X}$  được đưa về tối ưu hàm:

$$L(X) = \frac{1}{2s} \sum_{n=1}^N \sum_{m:r_{mn}=1} (\mathbf{y}_{mn} - \mathbf{x}_m \mathbf{w}_n)^2 + \frac{\lambda}{2} \|\mathbf{X}\|_F^2 \quad (3)$$

Để tối ưu hai hàm trên, ta sẽ sử dụng *Gradient Descent*.

Dễ thấy rằng, hàm (2) có thể tách thành  $\mathbf{N}$  bài toán nhỏ, mỗi bài toán ứng với việc đi tối ưu một cột của ma trận  $\mathbf{W}$ :

$$L(\mathbf{w}_n) = \frac{1}{2s} \sum_{m:r_{mn}=1} (\mathbf{y}_{mn} - \mathbf{x}_m \mathbf{w}_n)^2 + \frac{\lambda}{2} \|\mathbf{w}_n\|_2^2 \quad (4)$$

Vì biểu thức trong dấu  $\sum$  chỉ phụ thuộc vào các items đã được rated bởi user đang xét, ta có thể đơn giản nó bằng cách đặt  $\widehat{\mathbf{X}}_n$  là ma trận được tạo bởi các hàng tương ứng với các items đã được rated đó, và  $\widehat{\mathbf{y}}_n$  là các ratings tương ứng. Khi đó:

$$L(\mathbf{w}_n) = \frac{1}{2s} \|\widehat{\mathbf{y}}_n - \widehat{\mathbf{X}}_n \mathbf{w}_n\|_2^2 + \frac{\lambda}{2} \|\mathbf{w}_n\|_2^2 \quad (5)$$

và đạo hàm của nó:

$$\frac{\partial L(\mathbf{w}_n)}{\partial \mathbf{w}_n} = -\frac{1}{s} \widehat{\mathbf{X}}_n^T (\widehat{\mathbf{y}}_n - \widehat{\mathbf{X}}_n \mathbf{w}_n) + \lambda \mathbf{w}_n \quad (6)$$

Vậy công thức cập nhật cho mỗi cột  $\mathbf{W}$  là:

$$\mathbf{w}_n = \mathbf{w}_n - \eta \left( -\frac{1}{s} \widehat{\mathbf{X}}_n^T (\widehat{\mathbf{y}}_n - \widehat{\mathbf{X}}_n \mathbf{w}_n) + \lambda \mathbf{w}_n \right) \quad (7)$$

Tương tự như thế, mỗi cột của  $\mathbf{X}$ , tức vector cho mỗi *item*, sẽ được tìm bằng cách tối ưu:

$$L(\mathbf{x}_m) = \frac{1}{2s} \sum_{n:r_{mn}=1} (\mathbf{y}_{mn} - \mathbf{x}_m \mathbf{w}_n)^2 + \frac{\lambda}{2} \|\mathbf{x}_m\|_2^2 \quad (8)$$

Đặt  $\widehat{\mathbf{W}}_m$  là ma trận được tạo bởi các cột tương ứng với các *users* đã rate cho *item* đó, và  $\widehat{\mathbf{y}}_n$  là các *ratings* tương ứng. Khi đó (8) trở thành:

$$L(\mathbf{x}_m) = \frac{1}{2s} \|\widehat{\mathbf{y}}_n - \mathbf{x}_m \widehat{\mathbf{W}}_m\|_2^2 + \frac{\lambda}{2} \|\mathbf{x}_m\|_2^2 \quad (9)$$

Tương tự như trên, công thức cập nhật cho mỗi hàng của  $\mathbf{X}$  sẽ có dạng:

$$\mathbf{x}_m = \mathbf{x}_m - \eta \left( -\frac{1}{s} (\widehat{\mathbf{y}}_n - \mathbf{x}_m \widehat{\mathbf{W}}_m) \widehat{\mathbf{W}}_m^T + \lambda \mathbf{x}_m \right) \quad (10)$$



## Phương pháp đánh giá

### ❖ Độ tin cậy của giải thuật

Có nhiều phương pháp khác nhau có thể được sử dụng để đánh giá giải thuật như: F-Measure, Area Under the ROC Curve (AUC),... mỗi phương pháp đánh giá sẽ thích hợp cho từng lĩnh vực cụ thể. Trong RS, độ đo lỗi RMSE (Root Mean Squared Error) là độ đo phổ biến mà cộng đồng người dùng trong lĩnh vực RS thường sử dụng. RMSE hay được dùng cho bài toán dự đoán xếp hạng (Rating Prediction). RMSE được xác định bằng các công thức sau:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - f_i)^2}$$

RMSE (Root Mean Square Error) là giá trị bình phương trung bình lỗi lấy căn thường xuyên được sử dụng trong đánh giá độ khớp của mô hình so với dữ liệu huấn luyện. Được tính bằng căn bậc hai của giá trị tuyệt đối của hệ số tương quan giữa giá trị thực và giá trị dự đoán. [6].

### ❖ Áp dụng kiểm nghiệm

Áp dụng kiểm nghiệm với hệ thống gợi ý với bộ dữ liệu. Có 10k xếp hạng (điểm số 0-5, rating 0 đồng nghĩa với việc người đó chưa rating cho địa điểm tương ứng) bởi 1000 người dùng trên 100 địa điểm trong các ngữ cảnh về thời gian, thời tiết, bạn đồng hành.

Ta sẽ thực hiện dự đoán cho toàn bộ dữ liệu đang có (tất cả ngữ cảnh), 80% dữ liệu sẽ được dùng để huấn luyện, 20% còn lại được dùng để đánh giá. Kiểm tra với người dùng với id bất kì (VD: ở đây ta lấy user có id 124), ta thu được kết quả sau:

itemid	actual	predicted
1	4	2.459183
2	4	4.067507
3	2	3.685106
4	1	3.449183
5	4	2.5242

6	1	3.714869
7	2	1.716474
8	4	2.248005
9	0	2.42607
10	2	3.836276
11	3	4.21908
12	2	4.02709
13	0	1
14	3	1
15	1	3.238909
16	0	2
17	4	5
18	4	3.023607
19	1	4.038148
20	1	3.437161
21	4	4.27058
22	2	3.879009

*Bảng 3.5 Kết quả kiểm nghiệm của hệ thống gợi ý*

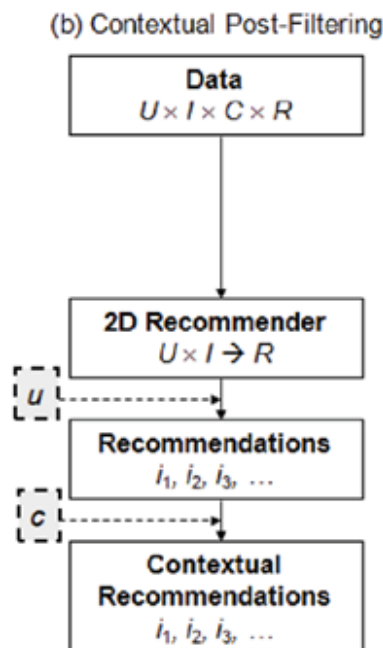
Ta sẽ thực hiện train dựa trên người dùng, ở đây, quá trình train lặp lại 100 lần, mỗi 10 lần ta sẽ in ra kết quả RMSE lúc đó, như ta thấy, giá trị hàm mất mát có chiều hướng giảm dần, giá trị RMSE có độ chênh lệch không đáng kể.

RMSE :1.5024

```
(env) C:\Users\AnT\Desktop\PY\MF>python run.py
run.py:10: FutureWarning: Method .as_matrix will be removed in a future version. Use .values instead.
  rate_train = ratings_base.as_matrix()
run.py:11: FutureWarning: Method .as_matrix will be removed in a future version. Use .values instead.
  rate_test = ratings_test.as_matrix()
iter = 10 , loss = 4.279200281940116 , RMSE train = 1.6126845911156258
iter = 20 , loss = 2.4815620501339817 , RMSE train = 1.4961509678728377
iter = 30 , loss = 1.7364923946983517 , RMSE train = 1.49134741290592
iter = 40 , loss = 1.3982476845965557 , RMSE train = 1.4912436663482138
iter = 50 , loss = 1.2432558650456331 , RMSE train = 1.4912645559961868
iter = 60 , loss = 1.1721723529065842 , RMSE train = 1.491271414381099
iter = 70 , loss = 1.1395697963800278 , RMSE train = 1.4912730872962652
iter = 80 , loss = 1.124616786056864 , RMSE train = 1.4912734748883045
iter = 90 , loss = 1.117758710880198 , RMSE train = 1.4912735633643912
iter = 100 , loss = 1.1146133115462462 , RMSE train = 1.4912735834119606
User-based MF, RMSE = 1.5024905789716356
```

Hình 3.5 Kết quả kiểm nghiệm

### 3.2. Phương pháp lọc ngữ cảnh sau



Hình 3.6 Phương pháp lọc ngữ cảnh sau

Khi sử dụng phương pháp lọc sau khi gợi ý, hệ thống sẽ áp dụng kỹ thuật tư vấn truyền thống để đưa ra những gợi ý. Sau đó sẽ tiến hành lọc loại bỏ những gợi ý mà không phù hợp với ngữ cảnh hiện tại của người dùng. Ví dụ hệ thống gợi ý địa điểm du lịch sẽ giới thiệu cho khách du lịch địa điểm gần họ nhất.

Để áp dụng kỹ thuật tư vấn truyền thống để đưa ra gợi ý, chúng ta cần chuyển đổi ma trận đánh giá nhiều chiều thành ma trận đánh giá hai chiều User-Item. Phương pháp lọc sau sẽ không tính toán độ tương tự dựa trên các tiêu chí của chiều ngữ cảnh. Do đó

ta sẽ chiếu toàn bộ đánh giá của người dùng  $u_j$  với item  $i_k$  lên mặt phẳng hai chiều. Các đánh giá khác nhau tại mỗi điểm ngữ cảnh sẽ được lấy thành giá trị trung bình cho đánh giá.

Giá trị của Rating được cho bởi công thức

$$R(u_j, i_k) = \frac{1}{r} \sum_{m=0}^p \sum_{n=0}^l R(u_j, i_k, C(m, n))$$

Trong đó

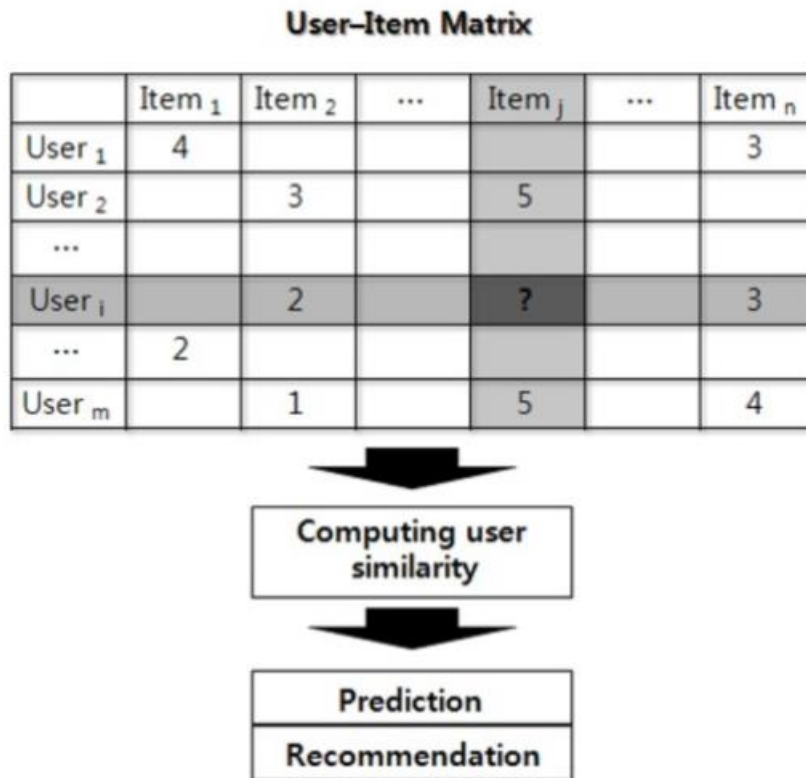
- $R(u_j, i_k)$  là giá trị đánh giá của  $u_j$  với item  $i_k$
- $p$  là số lượng chiều ngữ cảnh được xét trong hệ thống
- $l$  là số lượng giá trị ngữ cảnh được xét trong chiều ngữ cảnh  $C(m)$
- $r$  là số lượng bản ghi của người dùng  $u_j$  với item  $i_k$  tại tất cả các điểm ngữ cảnh
- $R(u_j, i_k, C(m, n))$  là giá trị đánh giá của  $u_j$  với item  $i_k$  tại điểm ngữ cảnh là  $C(m, n)$

Dữ liệu Bảng 3.1 của người dùng  $u1$  và item  $i1$  sau khi thực hiện chiếu lên mặt phẳng và tính lại giá trị rating. Dữ liệu được thể hiện như bảng bên dưới.

UserId	ItemId	Rating
u1	i1	2.63

*Bảng 3.6 Bảng user-item sau khi chiếu lên mặt phẳng*

Sau khi xử lý dữ liệu đánh giá. Kết quả ta thu được là rating của user-item được thể hiện bằng ma trận hai chiều. Khi đó ta sử dụng các phương pháp gợi ý truyền thống để thực hiện dự đoán và đưa ra gợi ý cho người dùng.



*Hình 3.7 Quá trình hoạt động của hệ thống gợi ý truyền thống*

Phương pháp lọc ngữ cảnh sau dùng hai yếu tố để tinh chỉnh kết quả nhằm đạt được hiệu quả tốt nhất là

- Lọc bỏ các gợi ý không liên quan đến ngữ cảnh đang áp dụng
- Hiệu chỉnh lại thứ tự xếp hạng các gợi ý trong danh sách.

### 3.3. Phân chia gợi ý theo ngữ cảnh

Một kỹ thuật phổ biến và có hiệu quả để đưa ra dự đoán cho hệ tư vấn có tích hợp yếu tố ngữ cảnh là phân chia sản phẩm (Item splitting). Khi đó, nhiều bản ghi của một địa điểm sẽ được tạo ra, mỗi bản ghi ứng với một yếu tố ngữ cảnh khác nhau trong không gian ngữ cảnh đang xét của hệ thống. Ví dụ với hệ thống đánh giá địa điểm du lịch nhận được nhiều đánh giá khác nhau của địa điểm đó vào ngày nắng và nhiều mây. Từ đó hệ thống sẽ tạo ra những bản sao khác nhau về địa điểm đó ngày nắng và ngày nhiều mây là  $i_s$  và  $i_c$ . Đánh giá vào ngày nắng sẽ được thể hiện là  $r_s$  và ngày nhiều mây sẽ là  $r_c$ .

Lợi ích của cách tiếp cận này là nó không đòi hỏi một thuật toán tư vấn mới, một thuật toán làm việc trên ma trận hai chiều ban đầu cũng có thể được áp dụng cho ma trận mới được chuyển đổi.

Để hiểu rõ hơn cách tiếp cận chia tách, ta xem xét ví dụ về đề xuất địa điểm theo bảng dữ liệu dưới đây

User	Item	Rating	Time	Weather	Companion
U1	I1	3	Weekend	Sunny	Partner
U1	I1	5	Weekend	Cloudy	Partner
U1	I1	?	Weekday	Cloudy	Family

*Bảng 3.7 Bảng dữ liệu đánh giá địa điểm trong ngữ cảnh*

Bảng dữ liệu trên bao gồm một người dùng U1, một mục sản phẩm I1, hai giá trị đánh giá đã biết và một giá trị đánh giá cần được dự đoán, ba yếu tố ngữ cảnh (Time(Weekend, Weekday); Location(Sunny, C); Companion(Partner, Family)).

Item splitting phân chia các sản phẩm theo những điều kiện ngữ cảnh. Việc phân chia nên được thực hiện khi thuật toán xác định được một điều kiện theo ngữ cảnh mà với từng giá trị của ngữ cảnh đó các sản phẩm được đánh giá khác biệt đáng kể. Trong ví dụ đánh giá các địa điểm du lịch ở trên, chiều ngữ cảnh Companion có 2 giá trị là: Partner, Family tương ứng có hai điều kiện thay “Partner và not Partner”, “Family và not Family”.

Item splitting lặp lại việc chia tách trên tất cả các điều kiện ngữ cảnh trong từng chiều ngữ cảnh và đánh giá hiệu quả của việc chia tách để lựa chọn điều kiện chia tách tốt nhất. Nó tìm kiếm cách chia tách tốt nhất cho mỗi sản phẩm để mỗi sản phẩm được chia tách thành hai loại mới và các ngữ cảnh được loại bỏ khỏi ma trận đánh giá ban đầu, ma trận đánh giá nhiều chiều trở thành ma trận đánh giá hai chiều. Giả sử rằng điều kiện ngữ cảnh tốt nhất để chia mục T1 trong Bảng 2.1 là “Weather = Sunny và not Sunny”, T1 có thể được chia thành T11 (địa điểm T1 lúc nắng) và T12 (địa điểm T1 không nắng). Ví dụ trên cho thấy một sự chia tách đơn giản, trong đó một điều kiện ngữ cảnh duy nhất được sử dụng để chia tách sản phẩm. Ngoài ra cũng có thể thực hiện việc phân chia phức tạp bằng cách sử dụng nhiều điều kiện trong nhiều ngữ cảnh.

## **Kết luận**

Chương này khóa luận đã trình bày các phương pháp để giải quyết bài toán đặt ra của hệ thống gợi ý, mỗi phương pháp đều có ưu điểm và nhược điểm.

Kỹ thuật phân rã ma trận đã được áp dụng thành công trong nhiều hệ thống, cách thực thi và cách thức để dự đoán ra kết quả phù hợp với người dùng hệ thống.

Đánh giá về độ chính xác của hệ thống gợi ý ngữ cảnh.

## **Chương 4: Giải pháp – Xây dựng ứng dụng**

Chương này thực hiện mô tả các thu thập dữ liệu sử dụng trong bài toán gợi ý có kết hợp yếu tố ngữ cảnh. Đồng thời đưa ra cách kiểm nghiệm tính chính xác của hệ thống. Và từng bước triển khai hệ thống gợi ý thành một hệ thống thực tế.

### **4.1. Tập dữ liệu**

Trong phạm vi khóa luận, tập dữ liệu được sử dụng để xây dựng hệ thống gợi ý là bộ dữ liệu Tourpedia. Là bộ dữ liệu về địa điểm du lịch được đánh giá ở nhiều các bối cảnh khác nhau.

#### **4.1.1. Giới thiệu về tập dữ liệu**

Tập dữ liệu Tourpedia được thu thập từ khảo sát yêu cầu của khách du lịch trên nhiều hình thức: dựa trên trải nghiệm người dùng, thông qua các bản khảo sát,..v..v. dựa vào đó, chọn ra các đánh giá, dữ liệu quan trọng, cần thiết để tổng hợp lại, phục vụ cho các mục đích du lịch của các hãng.

Số lượng dữ liệu (các bản ghi lại đánh giá) rất nhiều, vì nó tổng hợp lại của tất cả các đánh giá về các địa điểm trên thế giới, vì thế, dựa trên những trường dữ liệu ta quan tâm, sẽ lọc ra được những bản ghi cần thiết. Trong khóa luận này, em sẽ lấy ra 10k bản ghi đánh giá (điểm số 1-5) bởi 100 người dùng trên 100 địa điểm du lịch tại Hà Nội, các đánh giá dựa trên các ngữ cảnh như thời gian, thời tiết, bạn đồng hành.

Ưu điểm của tập dữ liệu này là các dữ liệu: tạo độ địa điểm, thông tin người dùng, địa điểm, đánh giá.. đều được tổng hợp qua API của google nên tính chính xác khá cao, và nếu cần thiết, ta có thể tự làm mới dữ liệu dựa trên các trường dữ liệu cũ:



	A	B	C	D	E	F
1	user_id	addr_id	rating	time	partner	weather
2	1	1	1	weekday	friends	cloudy
3	1	25	5	weekday	friends	rainy
4	1	66	3	holiday	friends	rainy
5	1	22	3	holiday	friends	sunny
6	1	67	1	holiday	alone	nice
7	1	36	1	weekend	alone	sunny
8	1	44	1	weekend	family	sunny
9	1	90	2	weekday	alone	nice
10	1	33	4	weekday	friends	sunny
11	1	48	4	weekday	alone	nice
12	2	70	3	holiday	friends	nice
13	2	12	4	holiday	family	sunny
14	2	93	4	holiday	alone	sunny
15	2	29	1	weekday	alone	cloudy
16	2	43	3	weekday	friends	cloudy
17	2	53	3	weekday	alone	nice
18	2	67	4	holiday	family	rainy
19	2	42	5	weekend	family	rainy
20	2	6	3	holiday	alone	rainy
21	2	16	1	weekday	alone	sunny
22	3	35	5	weekday	alone	nice
23	3	95	5	weekday	friends	nice
24	3	51	4	weekday	family	rainy
25	3	50	3	weekend	family	sunny
26	3	56	2	weekend	family	sunny
27	3	97	2	holiday	alone	nice
28	3	96	3	weekday	friends	rainy
29	3	34	4	weekday	friends	sunny
30	3	38	4	weekday	friends	sunny
31	3	10	5	weekday	alone	nice
32	4	33	4	holiday	friends	cloudy
33	4	45	2	weekday	family	sunny
34	4	15	5	weekend	alone	sunny
35	4	2	3	weekend	family	rainy
36	4	44	2	weekday	friends	rainy
37	4	40	2	holiday	alone	rainy

Hình 4.1 Bộ dữ liệu Tourpedia, kết hợp với Google

Cấu trúc mỗi bản ghi trong tập dữ liệu dùng trong hệ tự vấn gồm 5 trường: user\_id, addr\_id, rating, time, partner, weather.

- User\_id: Một số nguyên, là ID của người dùng được gán khi tham gia khảo sát.
- Addr\_id: Một số nguyên, là ID của địa điểm trên Cơ sở dữ liệu của Tourpedia.
- Rating: Là giá trị Rating của người dùng đó với địa điểm.
- Time: Là chuỗi ký tự mô tả thời gian thực hiện đánh giá bao gồm các giá trị như (Weekend, Weekday, Holiday)
- Weather: Là chuỗi ký tự mô tả về thời tiết của du khách khi đi du lịch tại các địa điểm đó (Sunny, nice, cloudy (ở đây ta sẽ quy chuẩn các điều kiện thời tiết khác thành 3 quy chuẩn trên)).

- Partner: Là chuỗi ký tự mô tả người đồng hành của du khách khi đi du lịch tại các địa điểm tương ứng (alone, family, friends)

#### 4.1.2. Phương pháp thu thập dữ liệu

Từ tập dữ liệu về các địa điểm download sẵn trên Tourpedia, ta sẽ lấy tọa độ của địa điểm, đánh giá dựa trên tập dữ liệu của Tourpedia.

Để làm sinh động thêm cho tập dữ liệu, ta sẽ kết hợp với dữ liệu của google API cho thêm gồm: thông tin người dùng, đánh giá, mô tả, ảnh địa điểm,.. bằng cách lấy tọa độ địa điểm ở tập dữ liệu của Tourpedia vào tra cứu bằng google map API.

A	B	C	D	E	F	G	H
STT	Address	place_id	types	lat	lng		
1	Cầu Long Biên, Cầu Long Biên, Bồ Đề, Long Biên, Hanoi	CHlhwov1WnTERKqWFMcwac	establishment,point_of_interest	21.0437359	105.85974		
2	Hồ Hoàn Kiếm, Hàng Trong, Hoàn Kiếm, Hanoi	CHlK0MSWnTERDqL5tGu_uG	establishment,natural_feature	21.0286669	105.8521484		
3	Trang Tien Ice Cream Company, Trang Tien, Hoàn Kiếm, Hanoi	CHlK47beuwrTERNv7pbUMN9Yw	establishment,food,point_of_interest,store	21.024867	105.85474		
4	Phố cổ Hà Nội, Đinh Tiên Hoàng, Lý Thái Tổ, Hoàn Kiếm, Hanoi	CHlK-LNyykNTERLUKx2QPY_M	establishment,point_of_interest,tourist_attraction	21.031569	105.8531015		
5	Bia hơi phố cổ Tạ Hiện	CHlK95CZ7-INTERgcmHMOEFw20	bar,establishment,food,point_of_interest,restaurant	21.0347898	105.8520742		
6	Phố ẩm thực Hàng Buồm	CHlKleLab-vTERIKYKIMavYPj0	establishment,food,point_of_interest,restaurant	21.0348594	105.8515334		
7	Phố sách Đinh Lễ	CHlKhdBa8SOINTER_c8c0k4e0e	establishment,point_of_interest,tourist_attraction	21.025998	105.8481242		
8	Phố Sách, 19 Tháng 12, Trần Hưng Đạo, Hoàn Kiếm, Hanoi	CHlKhdBa8SOINTER_c8c0k4e0e	establishment,point_of_interest,tourist_attraction	21.025998	105.8481242		
9	Nhà thờ Lớn Hà Nội, Nhà Chung, Hàng Trong, Hoàn Kiếm, Hanoi	CHlK57GypZWnTERz2JlTPBTT	church,establishment,place_of_worship,point_of_interest,tourist_attraction	21.0286657	105.8488123		
10	Chợ Đồng Xuân, Đồng Xuân, Hoàn Kiếm, Hanoi	CHlK68_-7rINTERN44F0D8aPw	political,sublocality,sublocality_level_1	21.0392122	105.8498353		
11	Vincom Center Ba Trĩu, Bà Triệu, Lê Đại Hành, Hai Bà Trưng District, Hanoi	CHlKPY2_Mw-sNTERnDv6wmp0DE	establishment,point_of_interest,shopping_mall	21.0114461	105.8500128		
12	Royal City, Ngõ 72 - Nguyễn Trãi, Khu đô thị Royal City, Thương Đình, Thanh Xuân, Hanoi	CHlK0QJP5usNTERmYKtoK0b-kk	establishment,point_of_interest,shopping_mall	21.000701	105.815971		
13	Times City, Phố Minh Khai, Vinhomes Times City, Vĩnh Tuy, Hai Bà Trưng District, Hanoi	CHlK8vOyA5sNTERqQJlTjvedMxA	establishment,point_of_interest,shopping_mall	20.9948261	105.8684447		
14	Vincom Center Nguyễn Chí Thanh, Nguyễn Chí Thanh, Láng Thượng, Đống Đa, Hanoi	CHlKXV4tGhNTERnAwM5egtc1Y	establishment,point_of_interest,shopping_mall	21.0233929	105.809093		
15	Vincom Phạm Ngọc Thạch, Phạm Ngọc Thạch, Kim Liên, Đống Đa, Hanoi	CHlK8GfUoerNTERnAwM5egtc1Y	establishment,point_of_interest,shopping_mall	21.0061882	105.8315717		
16	Vincom Long Biên, Bàng Láng, Vinhomes Riverside, Việt Hưng, Long Biên, Hanoi	CHlKfHAR6apNTERKxw_Kb29ak	route	21.049739	105.9110286		
17	vincom plaza bắc từ liêm	CHlK0eUjw9iUNTERgczTPRxtTSTl	clothing_store,establishment,point_of_interest,store	21.05268	105.7808137		
18	Hồ Tây, Tây Hồ, Hanoi	CHlKwZgcvcqNTER26QcYOYLeW	establishment,natural_feature	21.053238	105.8260943		
19	Nhà hát Lớn Hà Nội, Trang Tien, Trang Tien, Hoàn Kiếm, Hanoi	CHlK0B-OurNTERK41QzQdyemQ	establishment,point_of_interest,tourist_attraction	21.024184	105.857857		
20	Lăng Bác, Hùng Vương, Điện Biên, Ba Đình, Hanoi	CHlF138XqGnTERTE3h8KFDml	establishment,general_contractor,point_of_interest,tourist_attraction	21.0367789	105.8346447		
21	Rạp chiếu phim Quốc Gia, Thái Hà, Láng Hạ, Hanoi, Thành Công Đồng Đa Hà Nội	CHlJ3-SIXWONTERP6Kngf4Gkw	bus_station,establishment,point_of_interest,transit_station	21.0161269	105.815793		
22	Hoàng Thành Thăng Long, Hoàng Diệu, Điện Biên, Ba Đình, Hanoi	CHlJ5KwdKONTERmYlY9mmBU	establishment,point_of_interest,tourist_attraction	21.034381	105.8401142		
23	Văn Miếu- Quốc Tử Giám, Quốc Tử Giám, Văn Miếu, Đống Đa, Hanoi	CHlJGUlKSmmNTERJhaaOP0zTdM	establishment,point_of_interest,transit_station	21.0273766	105.8353074		
24	Cột cờ Hà Nội, Điện Biên Phủ, Điện Biên, Ba Đình, Hanoi	CHlKzVbuqNTER4I2SRfbkg	establishment,point_of_interest,tourist_attraction	21.0325556	105.8398039		
25	Hoa Lo Prison, Hoà Lò, Trần Hưng Đạo, Hoàn Kiếm, Hanoi	CHlJd5RqarNTERVK87gwhAZc	establishment,point_of_interest,tourist_attraction	21.025235	105.8463895		
26	Công viên Lê nin - Trần Phú, Trần Phú, Điện Biên, Ba Đình, Hanoi	CHlJGa6aDZnNTERbD0BQIAKVA	establishment,point_of_interest,transit_station	21.0310621	105.8394152		
27	Phủ Tây Hồ, Xóm Chua, Quang An, Tây Hồ, Hanoi	CHlJueTufZqNTERmjmkZRDyik	establishment,place_of_worship,point_of_interest,tourist_attraction	21.0551157	105.8197989		
28	Cầu Nhật Tân, Cầu Nhật Tân, Vĩnh Ngọc, Đông Anh, Hanoi	CHlK0cHsC4qNTERWYqra_UvoDE	route	21.0986244	105.8233477		
29	Làng hoa Ngọc Hà	CHlJOfv9vmmNTERPtnGk0DEpl	political,sublocality,sublocality_level_1	21.0378157	105.8282959		
30	Vườn Hoa Đào Lăng Nghè Nhật Tân, Ngõ 264 Âu Cơ, thôn Đồng, Nhật Tân, Tây Hồ, Hanoi	CHlJE-2C8v-qNTERl8D0l0w5GQ	establishment,park,point_of_interest,tourist_attraction	21.0738503	105.8273332		
31	Làng hoa Tây Tựu	CHlKpSQ5wRVNDEr_vh24DjPsv8	establishment,park,point_of_interest,tourist_attraction	21.0725029	105.7427672		
32	Làng hoa Nghi Tam	CHlK87CV15qNTERPV-flu6iUm	establishment,place_of_worship,point_of_interest	21.0565926	105.832831		
33	Làng hoa Mê Linh	CHlJLU471n-NDERm74Z9u5LZl	establishment,food,point_of_interest,restaurant	21.1695215	105.7399338		
34	Làng hoa Phú Thượng	CHlJgwDTSqNTER98-XDWHU_F4	political,sublocality,sublocality_level_1	21.0838905	105.8084823		
35	Làng hoa Liên Mạc	CHlKREPQD39IDER3gpp-SKsvE4	establishment, "point_of_interest", "school", "secondary_school"	21.1911106	105.6659158		
36	Làng hoa Từ Liêm	CHlJtFkoCqNTERMBON8R2RYX8	establishment,park,point_of_interest	21.0657694	105.8361064		

Hình 4.2 Dữ liệu địa điểm thu thập được bằng google map api

```
{
  "html_attributions": [],
  "result": {
    "name": "Google Australia",
    "reviews": [
      {
        "author_name": "Simon Cookes",
        "author_url": "https://www.google.com/maps/contrib/1173242463638831456/reviews",
        "language": "en",
        "profile_photo_url": "https://lh3.ggpht.com/-3h2uqJn47s/AAAAAAAAAII/AAAAAAAAAAQ/du71c6T0C/s128-cb00000000-cc-rp-mo/photo.jpg",
        "rating": 2,
        "relative_time_description": "trung tuần trước",
        "text": "Do not use Google Drive - cannot move between content manager folders even though it says you can, cannot share folders with external clients or partners, if you did not create the file then you cant move it the author of the file needs to do that so takes days to move hundreds of files as you need to copy them individually. Stick with Dropbox. Oh and unlimited space is not actually unlimited. (I would have given 1 star but know some great folks working at google)",
        "time": 1579456231
      },
      {
        "author_name": "Guru s",
        "author_url": "https://www.google.com/maps/contrib/106789570080138238260/reviews",
        "language": "en",
        "profile_photo_url": "https://lh3.ggpht.com/-DTR0Hva08Au/AAAAAAAAAII/AAAAAAAAAAQ/x1kETZK8Ij/s128-cb00000000-cc-rp-mo/photo.jpg",
        "rating": 5,
        "relative_time_description": "một tháng trước",
        "text": "Its been more than 3 years since I got the PR with great guidance & professional work from Dr. Anna. I had approached him this week for getting the passport details update . No questions asked. And the update was done in a most professional way. \n\nIf you want to get your feedback on whether you are eligible for PR , stop anything else you are doing and Dr. Anna is the person you have to speak to.\n\nBe ready to get the honest feedback on your status.",
        "time": 1578081387
      },
      {
        "author_name": "Bastian H Goonewardena",
        "author_url": "https://www.google.com/maps/contrib/116469171974372645286/reviews",
        "language": "en",
        "profile_photo_url": "https://lh3.ggpht.com/-zuVTGpU_4E/AAAAAAAAAII/AAAAAAAAAAQ/0pQmUFI0P-g/s128-cb00000000-cc-rp-mo-ba5/photo.jpg",
        "rating": 5,
        "relative_time_description": "2 tháng trước",
        "text": "It was a honour for me to get in to google Australia office. The staff is extremely helpful very kind. I had high expectations even before entering the office but the architecture and the design of the office itself blew my expectations over the roof. \n\nOne thing I noticed is that google actually cares about their staff and their well-being.\n\nI was there at a google education conference. The food that they serve was very good, possibly the best ever food I have had in a conference.",
        "time": 1568879793
      },
      {
        "author_name": "Mark Tittle",
        "author_url": "https://www.google.com/maps/contrib/106973681844973256152/reviews",
        "language": "en",
        "profile_photo_url": "https://lh3.ggpht.com/-0ej71L4jRSE/AAAAAAAAAII/AAAAAAAAAAQ/EBH9VUgaxZa/s128-cb00000000-cc-rp-mo-ba6/photo.jpg",
        "rating": 5,
        "relative_time_description": "một tháng trước",
        "text": "Great office, plenty of facilities and resources. So much respect for what Google does and their mission.",
        "time": 1571561285
      },
      {
        "author_name": "Budiono Sukses",
        "author_url": "https://www.google.com/maps/contrib/11118871451524069837/reviews",
        "language": "id",
        "profile_photo_url": "https://lh3.ggpht.com/-8eqD669L8zg/AAAAAAAAAII/AAAAAAAAAAQ/o_Ty1U0QkUx/s128-cb00000000-cc-rp-mo-ba7/photo.jpg",
        "rating": 5,
        "relative_time_description": "2 tháng trước",
        "text": "Kantor yang keren, besar, bersih, dengan lingkungan yang nyaman dan orang-orang yang bersahabat. Dekat dengan pelabuhan dan taman yang hijau dan indah. Datang ke sini untuk mengikuti Google Product Expert Meetup 2019",
        "time": 1566972047
      }
    ]
  },
  "status": "OK"
}
```

Hình 4.3 Dạng JSON đánh giá của người dùng

Dựa trên các dữ liệu thu thập được, ta sẽ tự làm mới, sinh ra, tích hợp lại thành các tập bản ghi để phục vụ cho mục đích sử dụng của khóa luận.

## 4.2. Tổng quan về hệ thống

### 4.2.1. Các thành phần chính của hệ thống

Các thành phần của hệ thống bao gồm:

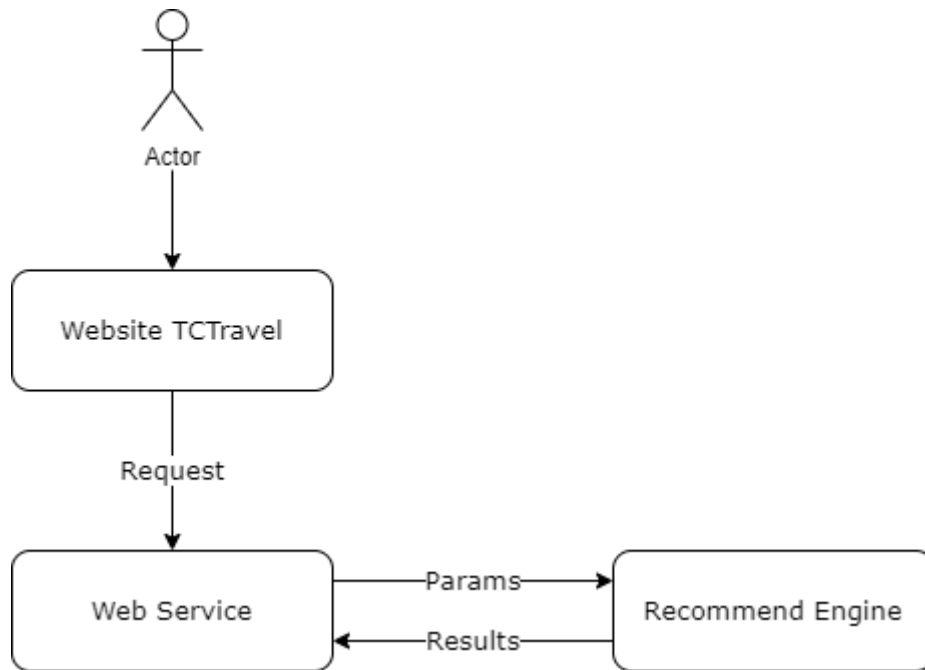
- **Website TCTravel**

Ứng dụng TCTravel là website tương tác với người dùng được xây dựng bằng framework Reactjs. Website là một landing page, vừa hỗ trợ người dùng với trang chia sẻ kinh nghiệm các địa điểm du lịch, vừa thêm thông tin (ảnh, video) về các địa điểm mà người dùng đã đến, vừa cung cấp giao diện tương tác giữa người dùng với hệ thống gợi ý.

- **Hệ thống gợi ý – Web service**

Hệ thống được xây dựng trên nền tảng nodejs xử lý toàn bộ yêu cầu truy vấn từ website và đưa ra dữ liệu api cho frontend xây dựng website.

### 4.2.2. Mô hình triển khai



Hình 4.4 Tổng quan về hệ thống

### 4.3. Yêu cầu hệ thống

Website cung cấp giao diện cho người dùng tham khảo kinh nghiệm du lịch tại các địa điểm hot. Thêm ảnh, video kỉ niệm về các địa điểm đã tham quan. Cho phép người dùng tra cứu địa điểm du lịch theo các thông tin mà người dùng cung cấp như: thời gian, thời tiết, bạn đồng hành, vị trí gần nhất,...v..v.

#### 4.3.1. Phân tích yêu cầu

Yêu cầu người dùng

- Giao diện dễ sử dụng, có tính thẩm mỹ cao, phù hợp với xu hướng thiết kế hiện tại.
- Cho phép khách hàng đăng nhập, đăng ký tài khoản.
- Xem kinh nghiệm du lịch từ những người đi trước.
- Xem lại, thêm mới bộ sưu tập về các địa điểm đã tới.
- Tra cứu hệ thống gợi ý du lịch

#### 4.3.2. Xác định tác nhân

Trong phạm vi của khóa luận, tác nhân duy nhất ta xét đến người dùng. Tác nhân quản lý thao tác với hệ thống gợi ý ở một hệ thống khác ở backend. Không sử dụng trực tiếp ứng dụng trên.

Người dùng:

- Đăng nhập, xem các kinh nghiệm về địa điểm du lịch.
- Xem, thêm bộ sưu tập về các địa điểm đã đi.
- Tra cứu hệ thống gợi ý du lịch

#### 4.3.3. Bảng thuật ngữ

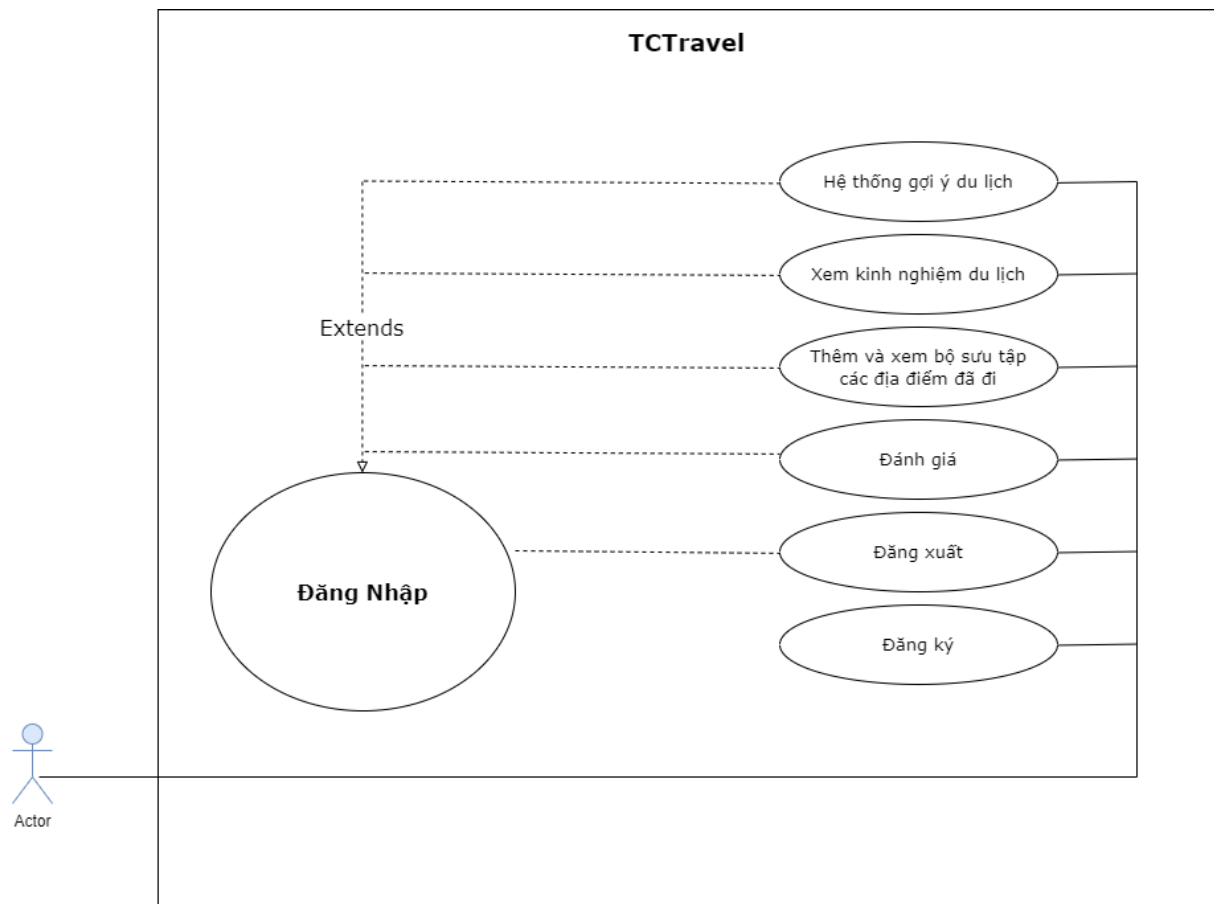
Việc xây dựng bảng thuật ngữ (Glossary) nhằm làm sáng tỏ các thuật ngữ được sử dụng cho một miền nào đó để mọi người hiểu, thống nhất được giữa các sản phẩm trong quá trình phát triển phần mềm.

STT	Tiếng Việt	Tiếng Anh	Giải thích
1	Người dùng	User	Người truy cập vào hệ thống, thực hiện xem thông tin địa điểm du lịch, v..v..
2	Sản phẩm (địa điểm du lịch)	Item	Là sản phẩm trong hệ thống gợi ý. Trong khóa luận đó là địa điểm du lịch.

*Bảng 4.1 Bảng thuật ngữ*

#### 4.3.4. Biểu đồ ca sử dụng – Usecase Diagram

Biểu đồ ca sử dụng (Usecase diagram) được dùng để chỉ ra quan hệ của các tác nhân với các ca sử dụng, cụ thể và quan hệ (relationship) giữa những ca sử dụng với nhau. Tuy nhiên, tác nhân duy trong phạm vi khóa luận, tác nhân duy nhất được xét đến là người dùng.



Hình 4.5 Use Case tổng quát của hệ thống

#### 4.3.5. Kịch bản

Kịch bản – Scenario miêu tả chi tiết quá trình hiện thực hóa hệ thống, giúp người phát triển hình dung rõ hơn về hệ thống.

Kịch bản của người dùng:

- Người dùng đăng nhập

<b>Tên Use Case</b>	<b>Đăng nhập</b>
<b>Tác nhân chính</b>	Người dùng
<b>Điều kiện trước</b>	Người dùng chưa đăng nhập
<b>Đảm bảo tối thiểu</b>	Hệ thống cho phép người dùng đăng nhập
<b>Điều kiện sau</b>	Nếu chức năng thành công, người dùng sẽ đăng nhập được vào hệ thống.  Ngược lại, trạng thái hệ thống không thay đổi.
<b>Chuỗi sự kiện chính:</b> <ol style="list-style-type: none"> <li>1. Người dùng chọn chức năng đăng nhập ở cuối menu trên đầu của website.</li> <li>2. Mở modal đăng nhập</li> <li>3. Người dùng nhập email và mật khẩu của mình</li> <li>4. Hệ thống kiểm tra tính hợp lệ của email và mật khẩu</li> <li>5. Đóng modal đăng nhập, người dùng tiếp tục sử dụng hệ thống</li> </ol>	
<b>Ngoại lệ:</b> <ol style="list-style-type: none"> <li>1. Người dùng nhập email và mật khẩu sai <ul style="list-style-type: none"> <li>➤ Hệ thống thông báo lỗi và yêu cầu nhập lại</li> </ul> </li> <li>2. Email người dùng đăng nhập không tồn tại <ul style="list-style-type: none"> <li>➤ Hệ thống thông báo lỗi. Yêu cầu tạo tài khoản mới</li> </ul> </li> </ol>	

*Bảng 4.2 Kịch bản người dùng đăng nhập*

- Người dùng đăng ký tài khoản

<b>Tên Use Case</b>	<b>Đăng ký</b>
<b>Tác nhân chính</b>	Người dùng
<b>Điều kiện trước</b>	Người dùng chưa có tài khoản
<b>Đảm bảo tối thiểu</b>	Hệ thống cho phép người dùng đăng ký
<b>Điều kiện sau</b>	Nếu chức năng thành công, người dùng sẽ đăng ký thành công tài khoản  Ngược lại, trạng thái hệ thống không thay đổi.
<b>Chuỗi sự kiện chính:</b> <ol style="list-style-type: none"> <li>1. Người dùng chọn chức năng đăng nhập ở cuối menu trên đầu của website.</li> <li>2. Bấm vào text “đăng ký” ở cuối modal dòng “Bạn chưa có tài khoản”</li> <li>3. Mở modal đăng ký</li> <li>4. Người dùng nhập thông tin theo yêu cầu</li> <li>5. Hệ thống kiểm tra tính hợp lệ của thông tin</li> <li>6. Đóng modal đăng ký, người dùng đăng nhập để sử dụng các chức năng</li> </ol>	
<b>Ngoại lệ:</b> <ol style="list-style-type: none"> <li>1. Người dùng nhập email đã tồn tại <ul style="list-style-type: none"> <li>➤ Hệ thống thông báo lỗi và yêu cầu nhập lại</li> </ul> </li> </ol>	

*Bảng 4.3 Kịch bản người dùng đăng ký*



- Người dùng tra cứu hệ thống gợi ý du lịch

Tên Use Case	Hệ thống gợi ý địa điểm du lịch
Tác nhân chính	Người dùng
Điều kiện trước	Người dùng đã đăng nhập vào hệ thống
Đảm bảo tối thiểu	Hệ thống báo lỗi.
Điều kiện sau	Nếu chức năng thành công, người dùng sẽ có thể sử dụng hệ thống gợi ý địa điểm du lịch.  Ngược lại, không thể sử dụng hệ thống.
<b>Chuỗi sự kiện chính:</b> <ol style="list-style-type: none"> <li>1. Người dùng truy cập website, chọn chức năng gợi ý trên menu.</li> <li>2. Cung cấp các thông tin để tra cứu: thời gian, thời tiết, bạn đồng hành, vị trí</li> <li>3. Hệ thống sẽ trả về các địa điểm gợi ý phù hợp trên website.</li> </ol>	
<b>Ngoại lệ:</b> <ol style="list-style-type: none"> <li>1. Ứng dụng không lấy được kết quả. <ul style="list-style-type: none"> <li>➤ Ứng dụng thông báo lỗi</li> </ul> </li> </ol>	

*Bảng 4.4 Kịch bản người dùng tra cứu hệ thống gợi ý*

- Xem kinh nghiệm du lịch

<b>Tên Use Case</b>	<b>Xem kinh nghiệm du lịch</b>
<b>Tác nhân chính</b>	Người dùng
<b>Điều kiện trước</b>	Người dùng đăng nhập vào hệ thống
<b>Đảm bảo tối thiểu</b>	Hệ thống báo lỗi truy cập
<b>Điều kiện sau</b>	Nếu chức năng thành công, người dùng sẽ thấy thông tin kinh nghiệm du lịch về các địa điểm nổi tiếng.  Ngược lại, hệ thống không hiển thị thông tin.
<b>Chuỗi sự kiện chính:</b> <ol style="list-style-type: none"> <li>1. Người dùng truy cập vào website.</li> <li>2. Chọn chức năng kinh nghiệm trên menu</li> <li>3. Hệ thống hiển thị giao diện chính có danh sách và tiêu đề của kinh nghiệm.</li> <li>4. Bấm vào để xem chi tiết kinh nghiệm.</li> </ol>	
<b>Ngoại lệ:</b> <ol style="list-style-type: none"> <li>1. Website không lấy được kết quả. <ul style="list-style-type: none"> <li>➤ Website thông báo lỗi</li> </ul> </li> </ol>	

*Bảng 4.5 Kịch bản người dùng xem kinh nghiệm người dùng*

- Thêm, xem bộ sưu tập

<b>Tên Use Case</b>	<b>Thêm, xem bộ sưu tập</b>
<b>Tác nhân chính</b>	Người dùng
<b>Điều kiện trước</b>	Người dùng đăng nhập vào hệ thống
<b>Đảm bảo tối thiểu</b>	Hệ thống báo lỗi truy cập
<b>Điều kiện sau</b>	Nếu chức năng thành công, người dùng sẽ thấy được bộ sưu tập và thêm mới bộ sưu tập.  Ngược lại, hệ thống không hiển thị thông tin.
<b>Chuỗi sự kiện chính:</b> <ol style="list-style-type: none"> <li>1. Người dùng truy cập vào website.</li> <li>2. Chọn chức năng bộ sưu tập trên menu</li> <li>3. Hệ thống hiển thị giao diện chính có button thêm mới và danh sách bộ sưu tập về các địa điểm đã đi.</li> <li>4. Bấm vào button thêm mới để sử dụng chức năng.</li> </ol>	
<b>Ngoại lệ:</b> <ol style="list-style-type: none"> <li>1. Website không lấy được kết quả. <ul style="list-style-type: none"> <li>➤ Website thông báo lỗi</li> </ul> </li> <li>2. Hệ thống không thêm mới được bộ sưu tập</li> </ol>	

*Bảng 4.6 Kịch bản người dùng thêm, xem bộ sưu tập*

- Đánh giá địa điểm du lịch

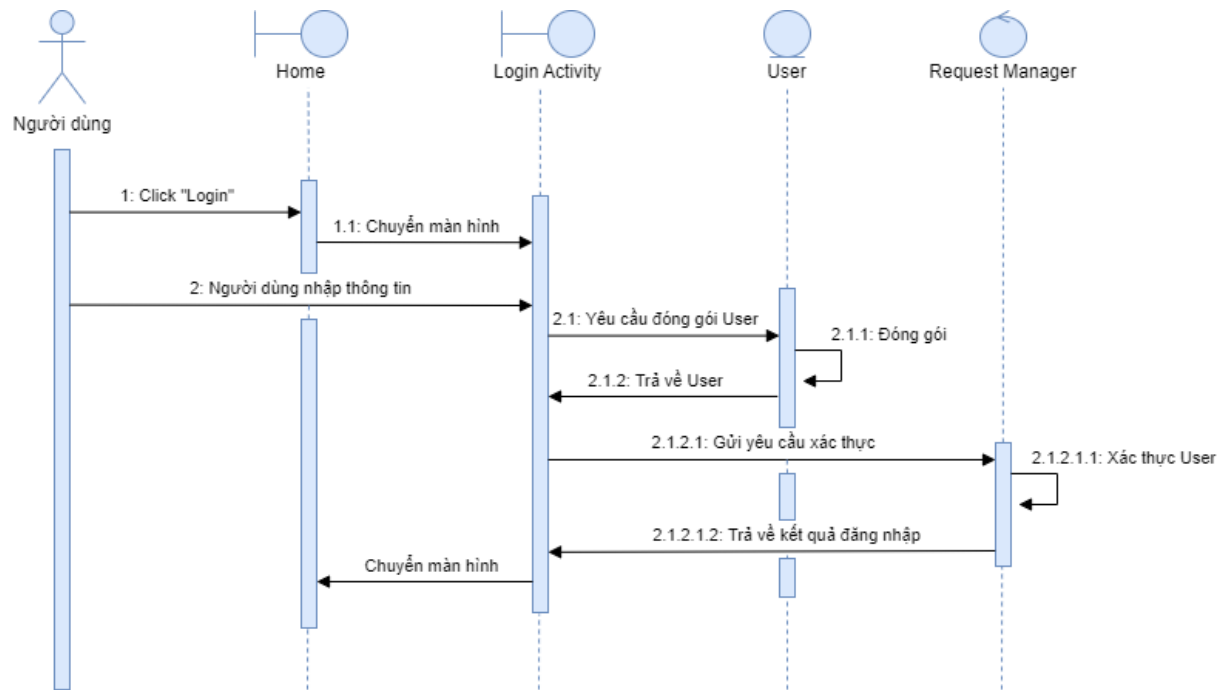
Tên Use Case	Đánh giá
Tác nhân chính	Người dùng
Điều kiện trước	Người dùng đã đăng nhập vào hệ thống
Đảm bảo tối thiểu	Hệ thống báo lỗi việc đánh giá
Điều kiện sau	Nếu chức năng thành công, người dùng sẽ thấy mức độ đánh giá của mình tương ứng với số lượng sao theo từng địa điểm.  Ngược lại, hệ thống sẽ báo lỗi.
<b>Chuỗi sự kiện chính:</b> <ol style="list-style-type: none"> <li>1. Người dùng truy cập vào website chọn chức năng gợi ý trên menu.</li> <li>2. Chọn action đánh giá ở bảng các địa điểm được gợi ý theo yêu cầu đã chọn.</li> <li>3. Hiện thị modal đánh giá.</li> <li>4. Người dùng nhập điểm đánh giá để đưa ra mức độ ưa thích.</li> <li>5. Hệ thống thực hiện gửi đi kết quả đánh giá của người dùng.</li> <li>6. Hệ thống hiện thị đánh giá theo sự lựa chọn của người dùng.</li> </ol>	
<b>Ngoại lệ:</b> <ol style="list-style-type: none"> <li>1. Hệ thống không thể gửi đi kết quả đánh giá của người dùng. <ul style="list-style-type: none"> <li>➤ Hệ thống thông báo lỗi việc xử lý.</li> </ul> </li> </ol>	

Bảng 4.7 Kịch bản người dùng đánh giá

## 4.4. Thiết kế hệ thống

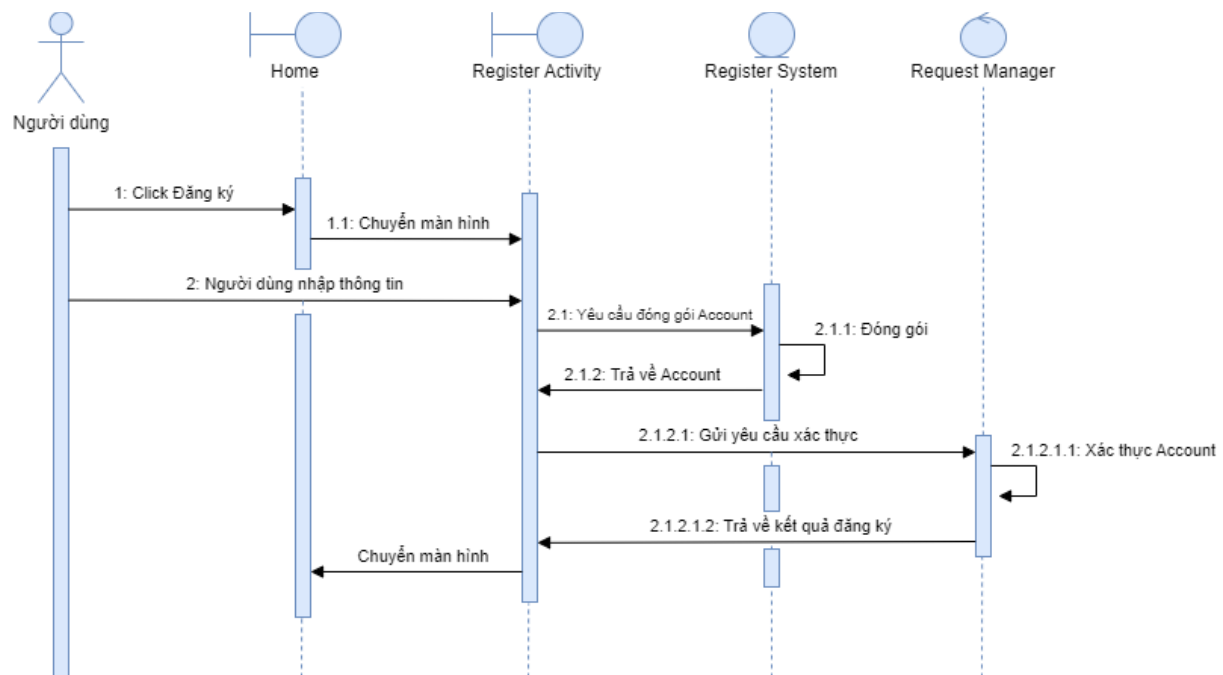
### 4.4.1. Sơ đồ tuần tự

- Người dùng đăng nhập



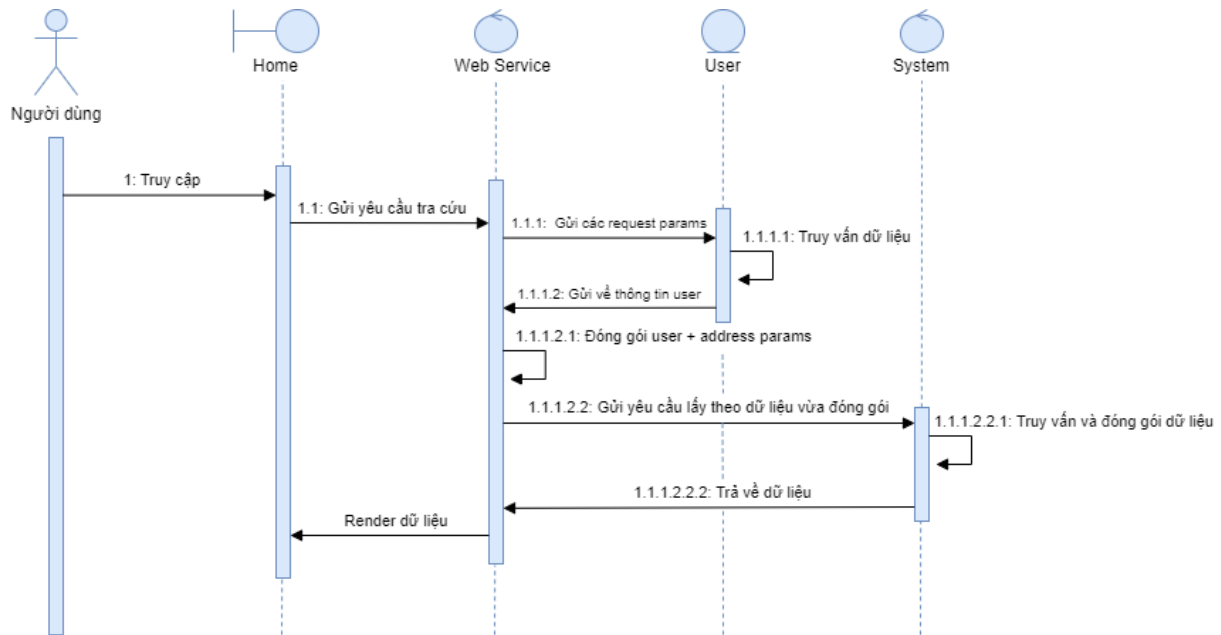
Hình 4.6 Biểu đồ tuần tự - Người dùng đăng nhập

- Người dùng đăng ký



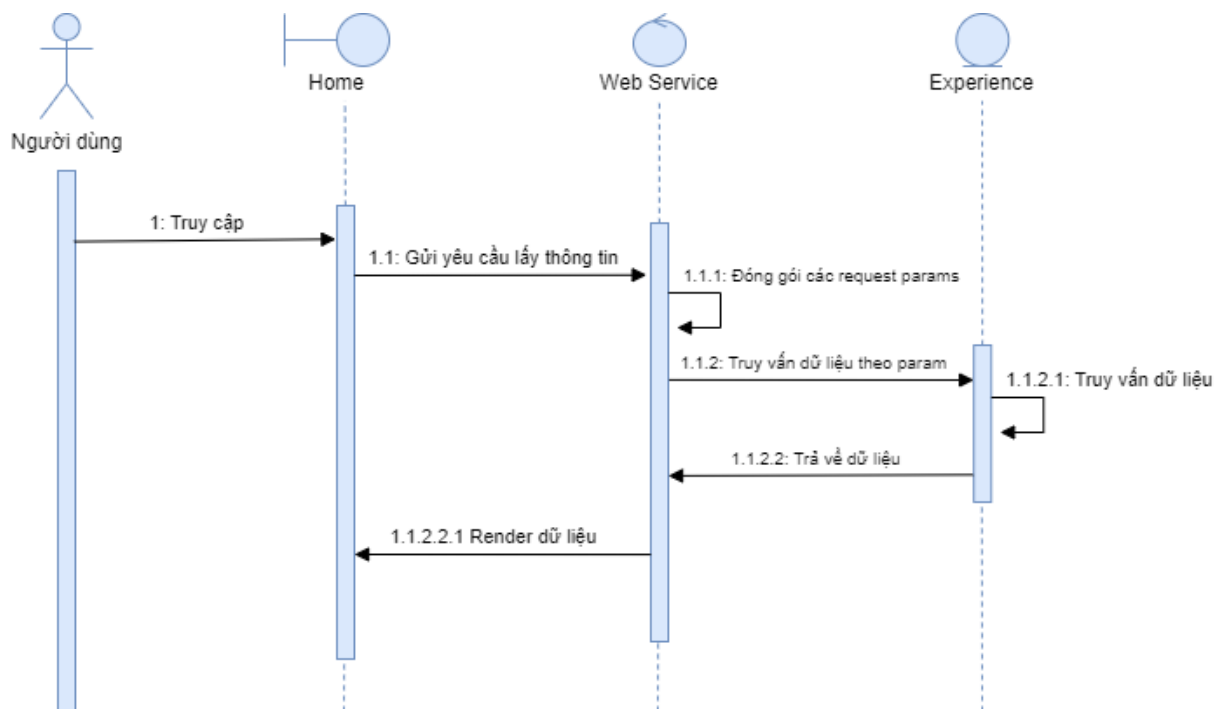
Hình 4.7 Biểu đồ tuần tự - Người dùng đăng ký

- Người dùng tra cứu hệ thống gợi ý du lịch



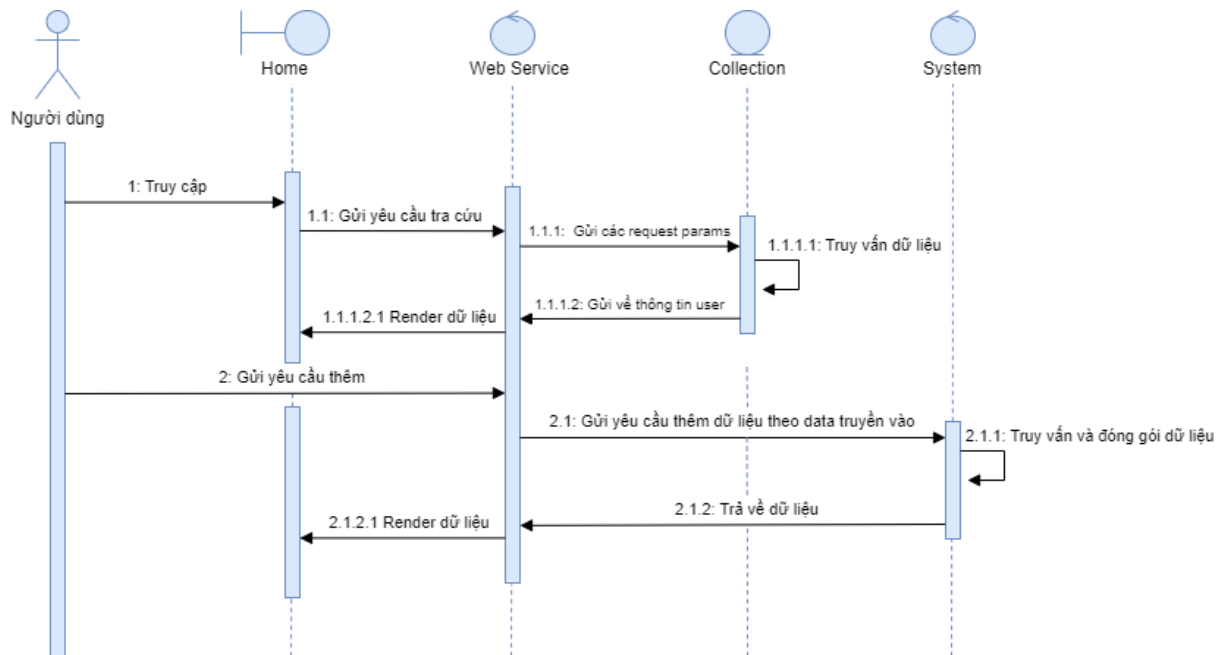
Hình 4.8 Biểu đồ tuần tự - Người dùng tra cứu hệ thống gợi ý du lịch

- Người dùng xem kinh nghiệm du lịch



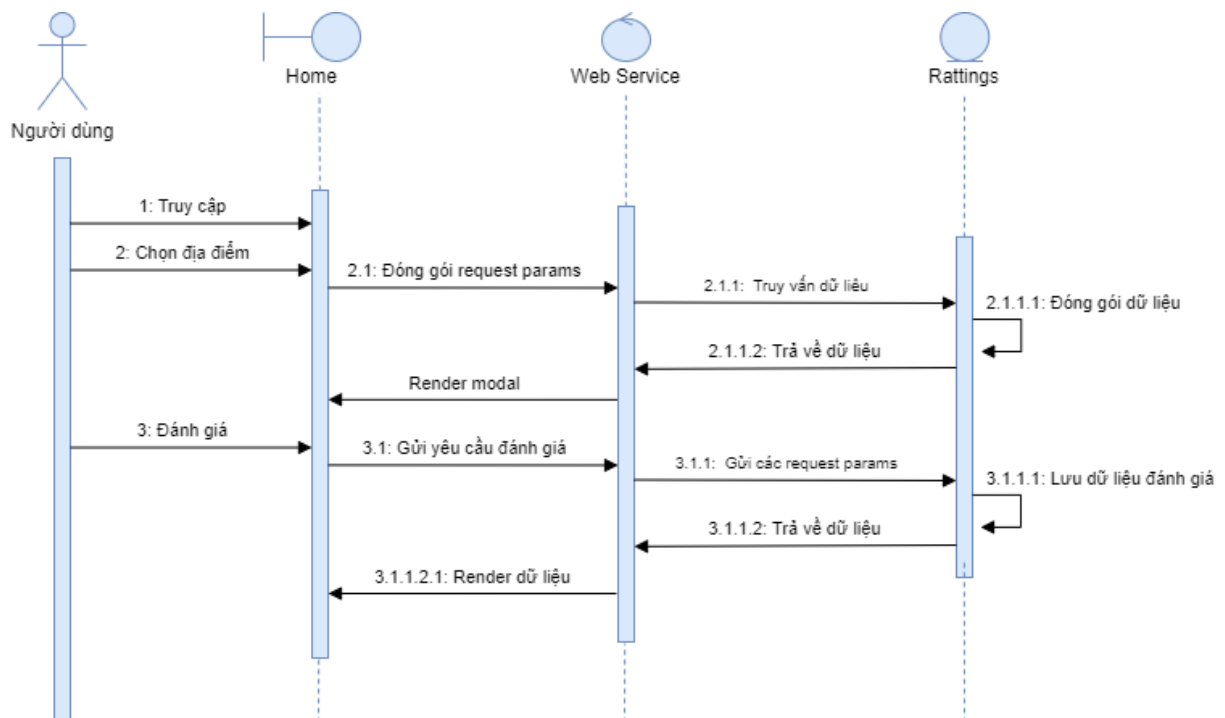
Hình 4.9 Biểu đồ tuần tự - Người dùng xem kinh nghiệm du lịch

- Người dùng thêm, xem bộ sưu tập



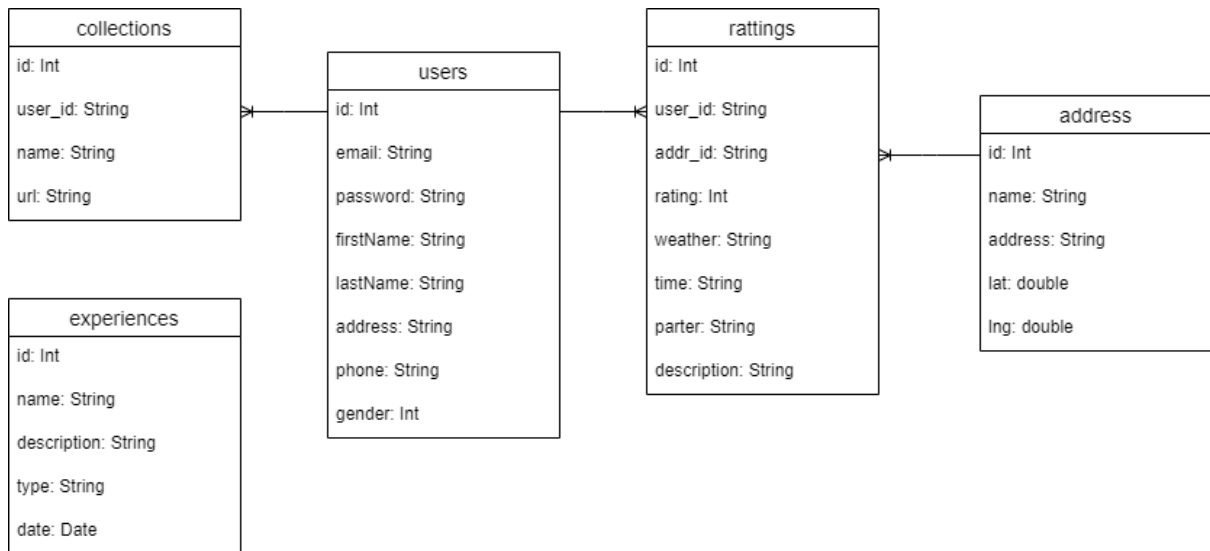
Hình 4.10 Biểu đồ tuần tự - Người dùng thêm, xem bộ sưu tập

- Người dùng đánh giá địa điểm



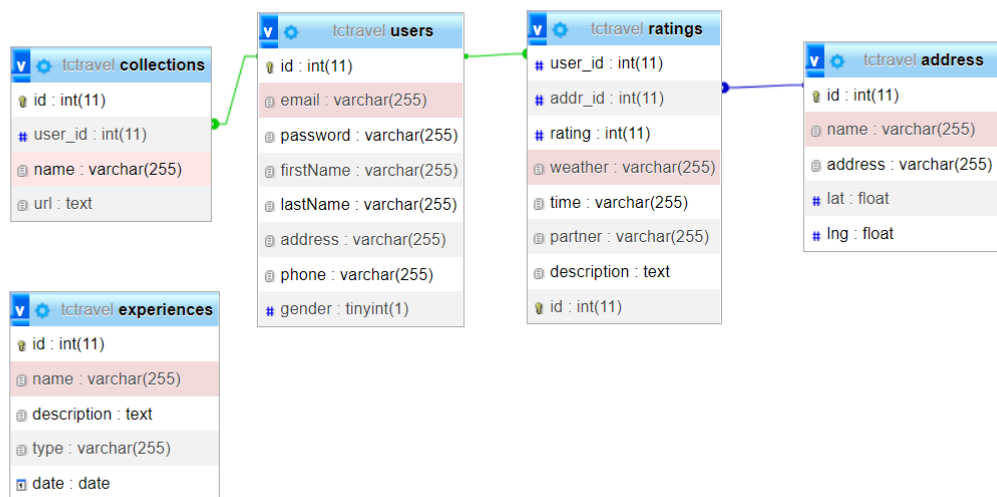
Hình 4.11 Biểu đồ tuần tự - Người dùng đánh giá địa điểm

#### 4.4.2. Biểu đồ thực thể thiết kế



Hình 4.12 Biểu đồ thực thể

#### 4.4.3. Lược đồ cơ sở dữ liệu

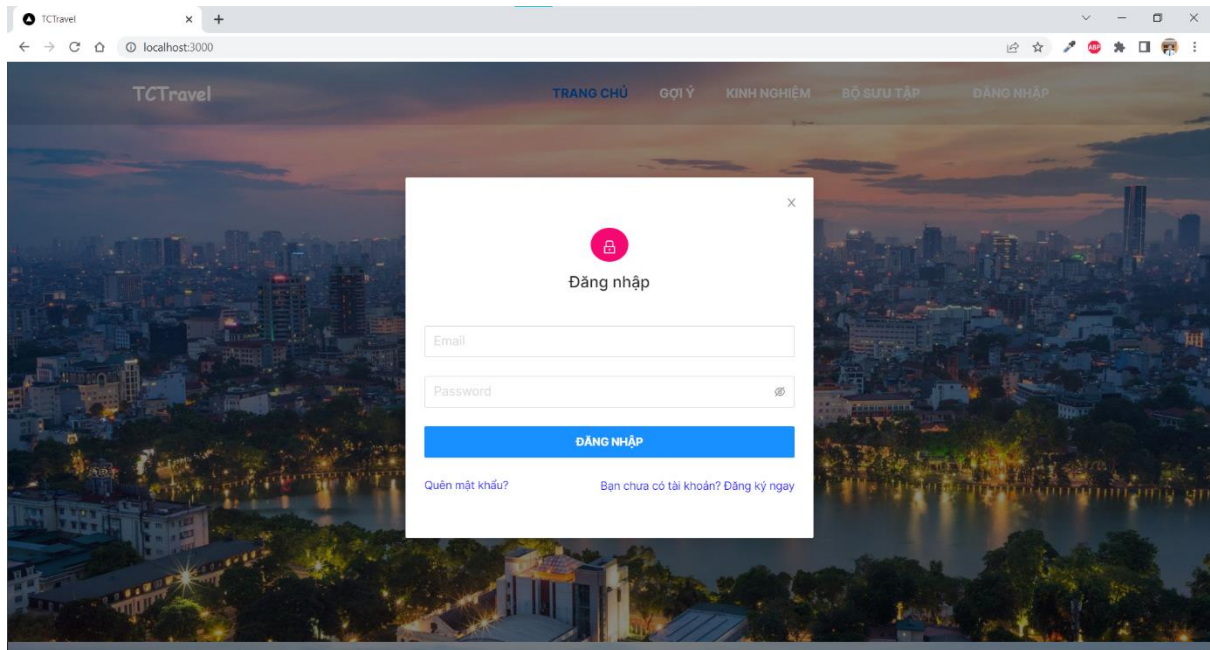


Hình 4.13 Lược đồ cơ sở dữ liệu

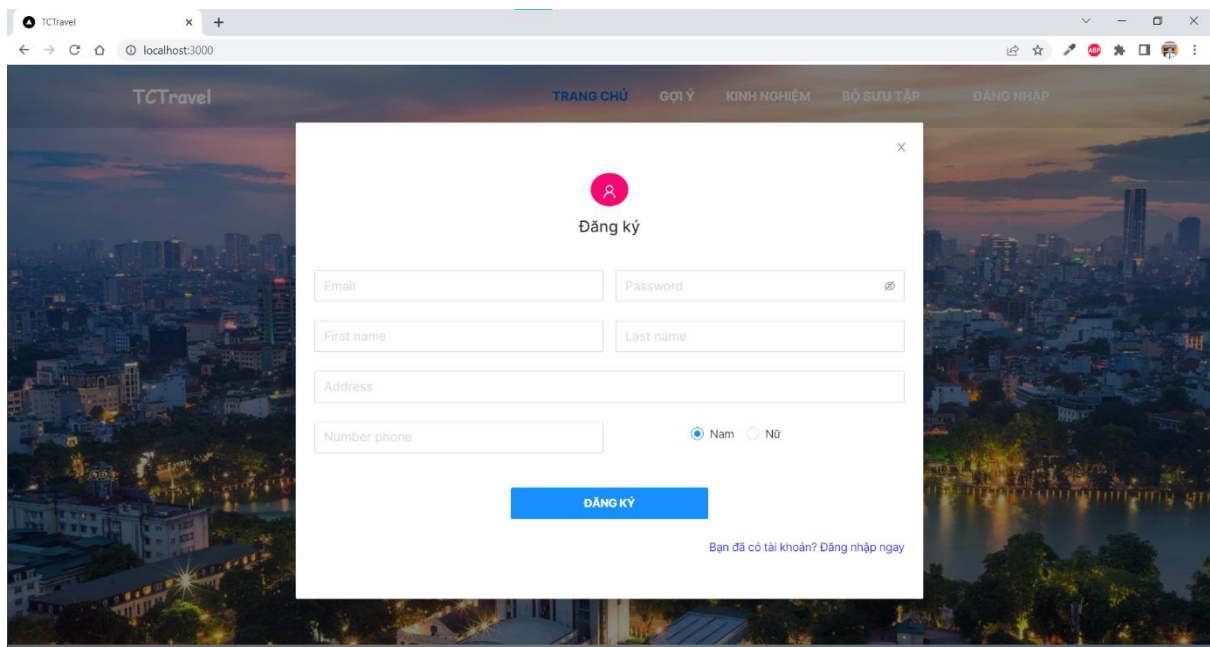


## 4.5. Kết quả

### \* Chức năng đăng nhập, đăng ký

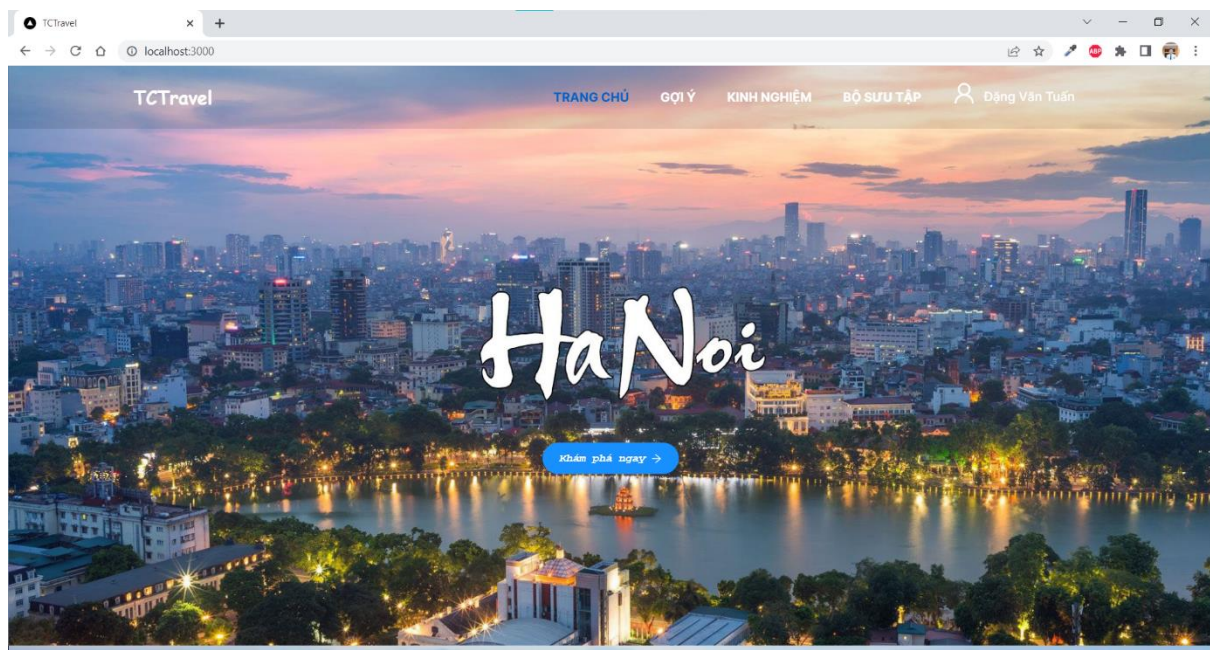


Hình 4.14 Giao diện màn hình đăng nhập



Hình 4.15 Giao diện màn hình đăng ký

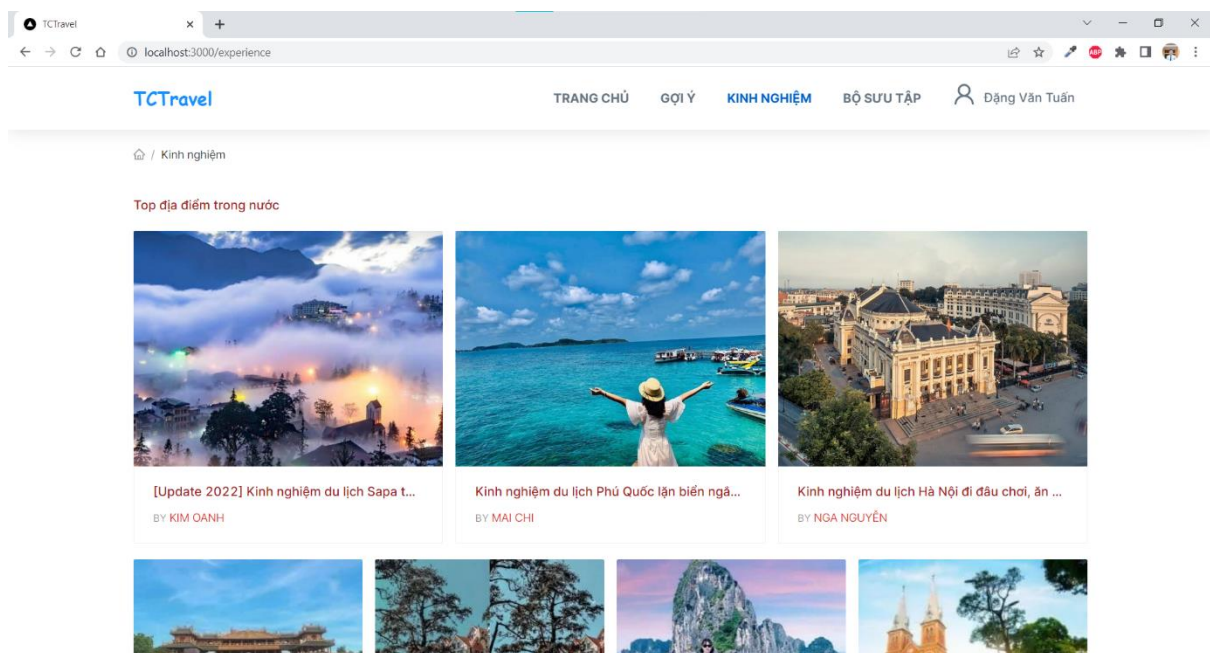
## \* Giao diện trang chủ



Hình 4.16 Giao diện màn hình trang chủ

Ở màn này người dùng có thể di chuyển giữa các màn hình chức năng ở thanh menu trên cùng. Ở góc bên phải hiển thị tên người dùng đã đăng nhập hoặc nút đăng nhập.

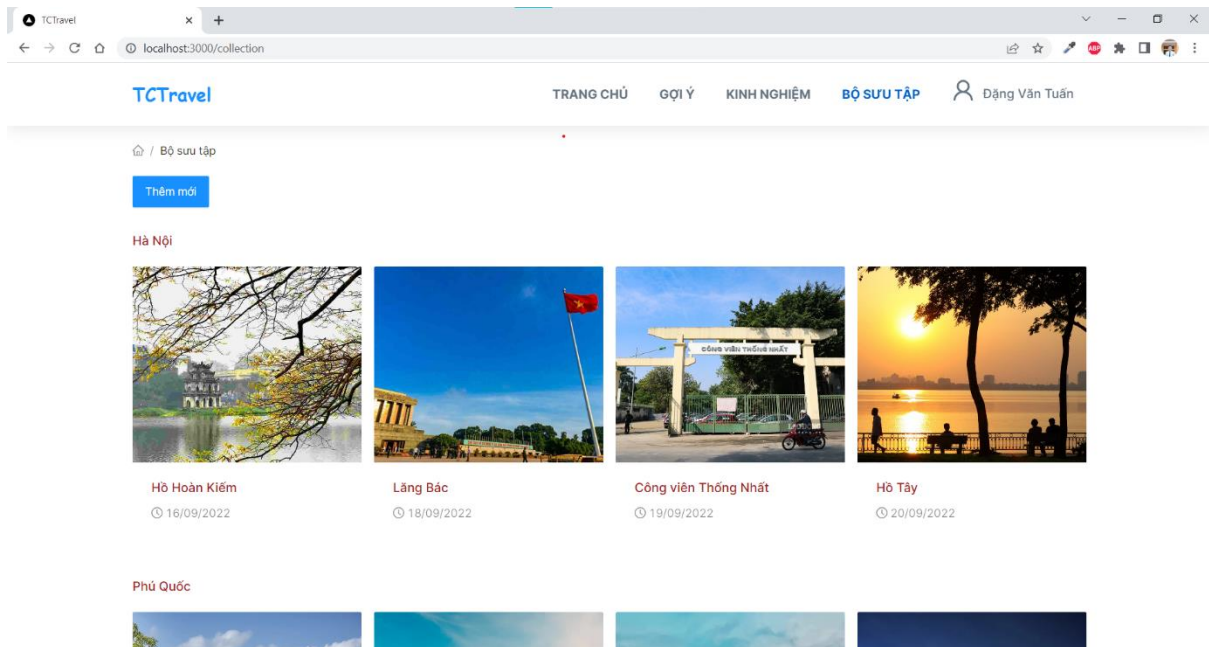
## \* Giao diện màn hình kinh nghiệm



Hình 4.17 Giao diện màn hình kinh nghiệm

Trong màn này người dùng có thể xem danh sách chia sẻ kinh nghiệm của các địa điểm nổi tiếng ở trong nước và ngoài nước. Người dùng bấm vào từng mục để xem chi tiết.

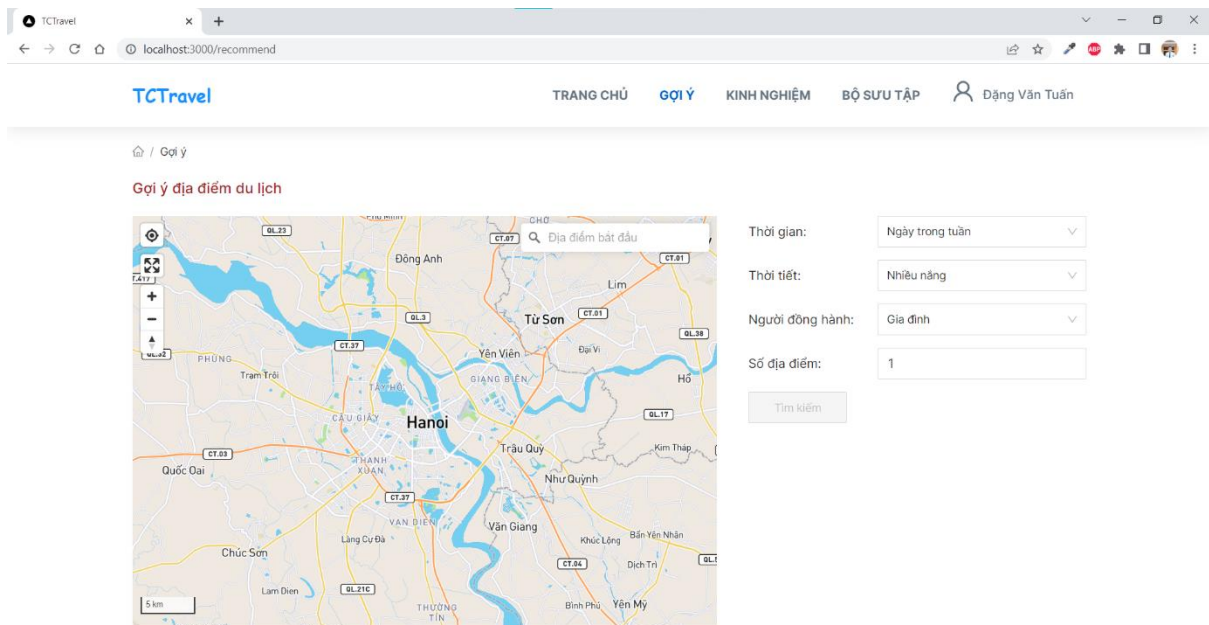
## \* Giao diện màn hình bộ sưu tập



Hình 4.18 Giao diện màn hình bộ sưu tập

Ở màn hình này người dùng có thể xem lại hình ảnh, video các địa điểm đã từng đến. Người dùng có thể thêm mới bộ sưu tập.

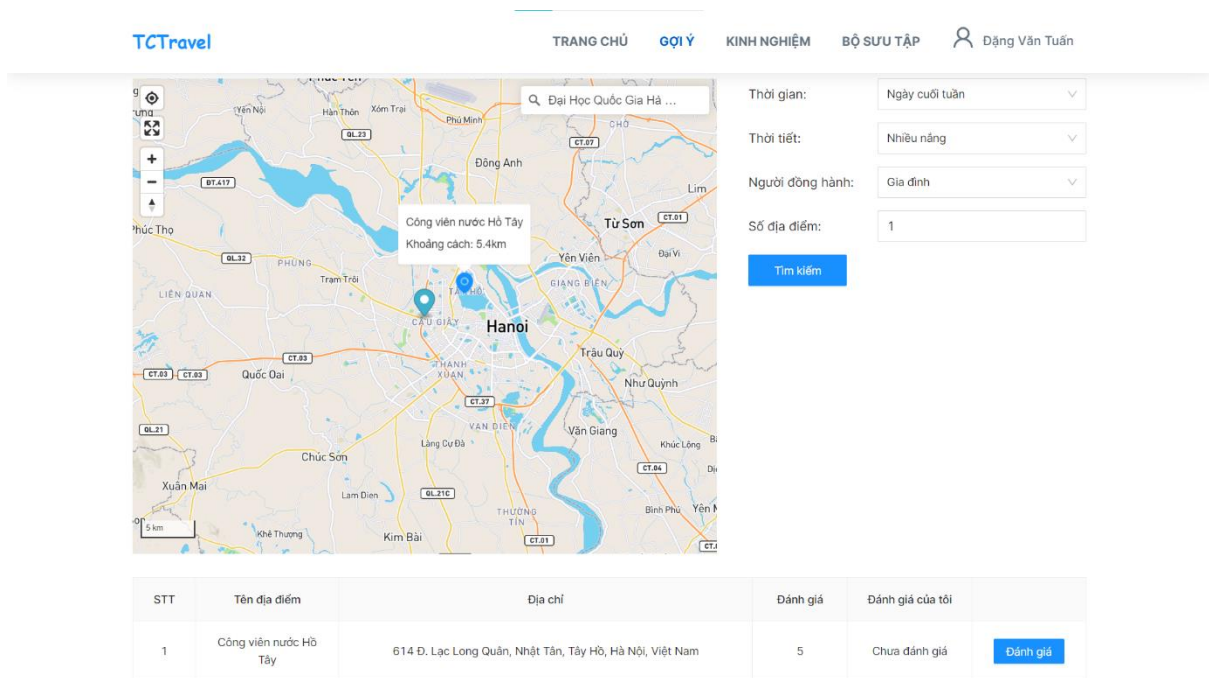
## \* Chức năng gợi ý



Hình 4.19 Màn hình gợi ý ban đầu

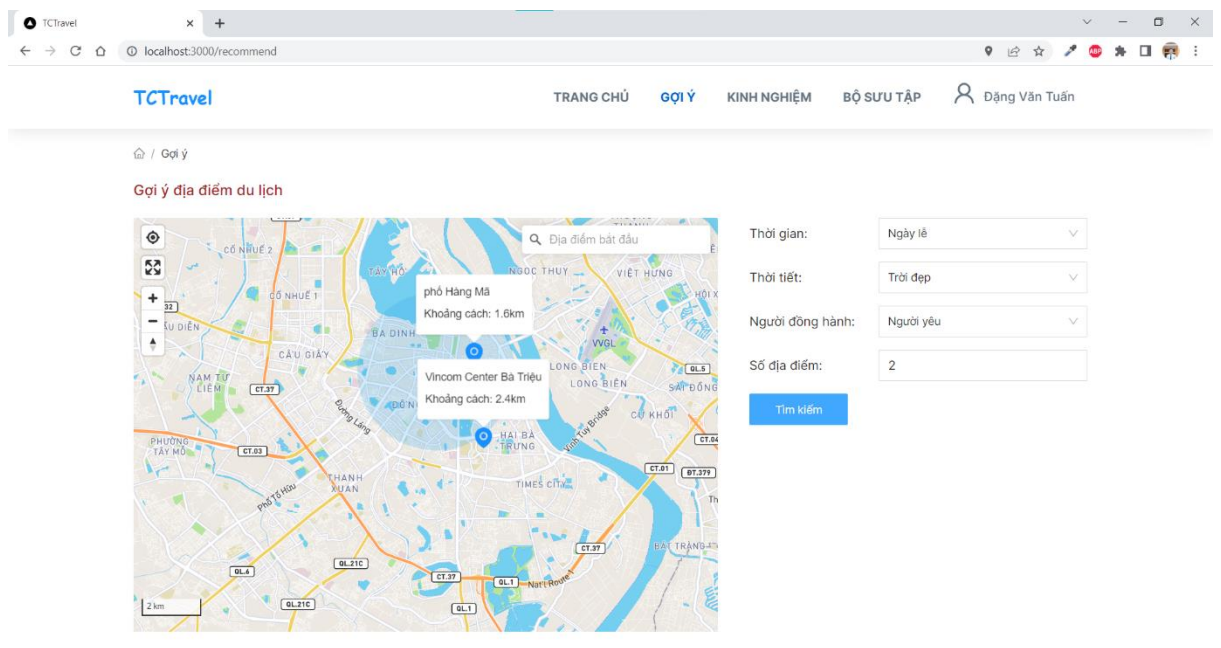
Đây là giao diện màn trang gợi ý ban đầu với map bên trái và các tùy chọn ngữ cảnh ở bên phải.





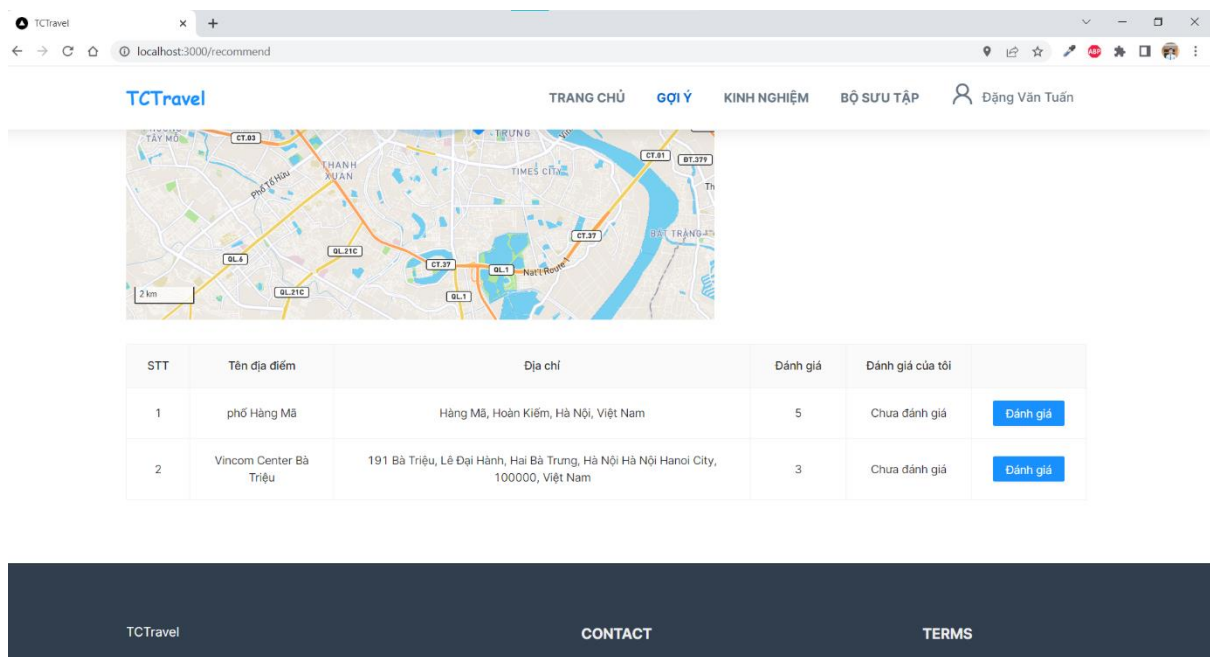
Hình 4.20 Màn hình gợi ý khi nhập vị trí xuất phát và kết quả

Giao diện màn hình kết quả khi người dùng nhập địa điểm xuất phát ở ô input địa điểm bắt đầu và chọn các ngữ cảnh gồm thời gian, thời tiết, người đồng hành cùng với số địa điểm người dùng mong muốn. Kết quả trả ra thông tin về địa điểm ở bảng dưới cùng, đánh dấu địa điểm ở trên map và khoảng cách từ điểm xuất phát tới điểm đó.

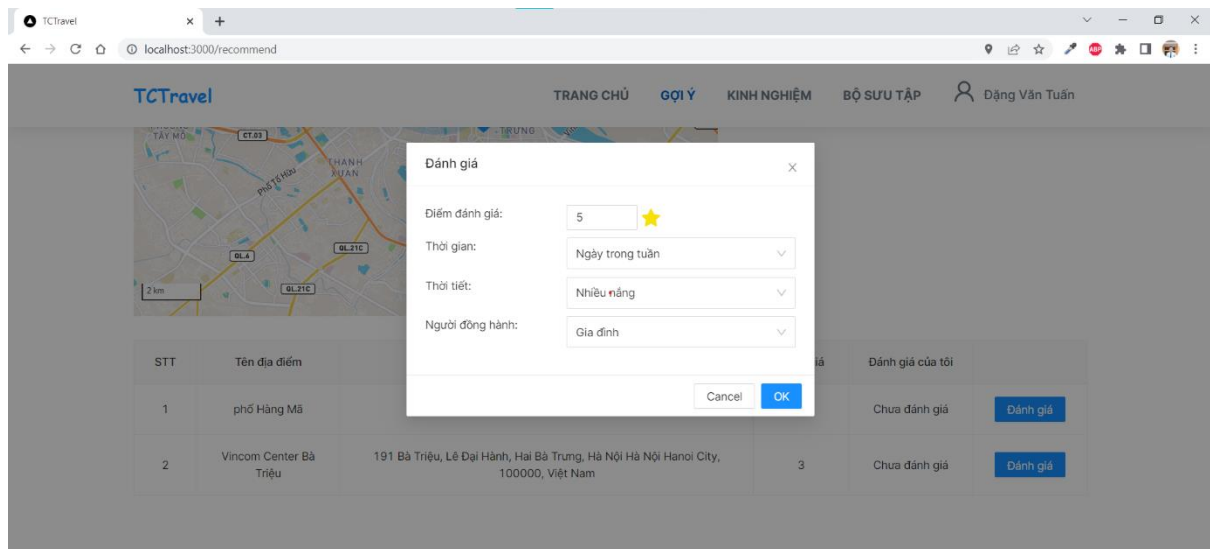


Hình 4.21 Màn hình gợi ý với điểm xuất phát là vị trí hiện tại

Giao diện màn hình khi chọn địa điểm xuất phát là vị trí hiện tại và thông tin ngữ cảnh.



Hình 4.22 Màn hình kết quả với điểm xuất phát là vị trí hiện tại



Hình 4.23 Màn hình đánh giá địa điểm

## **Chương 5 Kết luận**

Nhu cầu tham quan du lịch, khám phá của con người ngày một tăng cao vì thế áp dụng công nghệ kỹ thuật vào đời sống để phục vụ nhu cầu đó là rất phù hợp và cần thiết. Khóa luận nhằm giải quyết mục đích đó đã xây dựng hệ thống đưa gợi ý cho người dùng về các địa điểm du lịch và hỗ trợ người dùng lưu trữ nhật ký của từng địa điểm.

### **Kết quả khóa luận thu được**

Khóa luận đã nghiên cứu và xây dựng thành công một hệ tư vấn cộng tác dựa trên ngữ cảnh. Khóa luận đã đưa ra các phương pháp tư vấn khác và nhận thấy phương pháp lọc trước ngữ cảnh trước, lọc sau và với kỹ thuật xử lý chính là phân rã ma trận, kết hợp với việc sàng lọc lại dữ liệu Tourpedia để giúp hệ gợi ý tốt hơn.

Thiết kế hoàn thiện hệ thống gợi ý địa điểm du lịch, cài đặt thành công ứng dụng trên local, thử nghiệm với dữ liệu giả định các địa điểm ở khu vực Hà Nội.

### **Hướng phát triển tương lai**

Xây dựng thêm tính năng hỗ trợ người dùng như chỉ đường giữa các địa điểm, xem chi tiết thông tin mô tả về các địa điểm, kinh nghiệm tham quan các địa điểm đó.

Mong muốn sử dụng bộ lõi xử lý thông tin này để áp dụng vào các hệ thống thương mại du lịch khác nhằm nâng cao doanh số cho các hệ thống. Đồng thời nghiên cứu và phát triển các phương pháp mới cho hệ tư vấn ngữ cảnh.

## TÀI LIỆU THAM KHẢO

- [1] “Dulich9.com”, [Online].Available: <https://dulich9.com>. [Accessed April 2019].
- [2] <https://blog.dominodatalab.com/recommender-systems-collaborative-filtering/>
- [3] [https://en.wikipedia.org/wiki/Cosine\\_similarity](https://en.wikipedia.org/wiki/Cosine_similarity)
- [4] [https://en.wikipedia.org/wiki/Pearson\\_correlation\\_coefficient](https://en.wikipedia.org/wiki/Pearson_correlation_coefficient)
- [5][https://medium.com/@m\\_n\\_malaeb/recall-and-precision-at-k-for-recommender-systems-618483226c54](https://medium.com/@m_n_malaeb/recall-and-precision-at-k-for-recommender-systems-618483226c54)
- [6] “PYS travel”, [Online].Available: <https://pystravel.vn>. [Accessed April 2019].
- [7] Kwan Hui Lim, Xiaoting Wang, Jeffrey Chan, Shanika Karunasekera, Christopher Leckie, Yehui Chen, Cheong Loong Tan, Fu Quan Gao, Teh Ken Wee, “PersTour: A Personalized Tour recommendation and Planning System”, Department of Computing and Information System, The University of Melbourne, Austraylia.
- [8] <https://visgl.github.io/react-map-gl/>
- [9] <https://docs.mapbox.com/mapbox-gl-js/example/mapbox-gl-geocoder/>
- [10] <https://docs.mapbox.com/mapbox-gl-js/example/>
- [11] <https://machinelearningcoban.com/2017/05/31/matrixfactorization/>