{Trang trắng này dùng để dán bản Nhận xét của người hướng dẫn, hoặc thay trang này bằng Nhận xét của người hướng dẫn}

{Trang trắng này dùng để dán bản Nhận xét của người phản biện, hoặc thay trang này bằng Nhận xét của người phản biện}

**TÓM TẮT**

Tên đề tài: Xây dựng ứng dụng gợi ý du lịch hướng ngữ cảnh

Sinh viên thực hiện: Tăng Văn Quốc Chí

Số thẻ SV: 102140060 Lớp: 14T2

………………………………………………………………………………………… ..

………………………………………………………………………………………… ..

………………………………………………………………………………………… ..

………………………………………………………………………………………… ..

………………………………………………………………………………………… ..

………………………………………………………………………………………… ..

………………………………………………………………………………………… ..

|  |  |
| --- | --- |
| ĐẠI HỌC ĐÀ NẴNG  **TRƯỜNG ĐẠI HỌC BÁCH KHOA**  KHOA ………………………………………… | **CỘNG HÒA XÃ HÔI CHỦ NGHĨA VIỆT NAM**  Độc lập - Tự do - Hạnh phúc |

**NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP**

Họ tên sinh viên: …..…………….………….…….. Số thẻ sinh viên: ………………...

Lớp:…………… Khoa:....................................... Ngành: ……………….......................

1. *Tên đề tài đồ án:*

………………………………………………..…………………………………………

…………………………………………………………………………………………..

1. *Đề tài thuộc diện:*  *Có ký kết thỏa thuận sở hữu trí tuệ đối với kết quả thực hiện*
2. *Các số liệu và dữ liệu ban đầu:*

……………………………………..……………………………………………..……......……………………………………………………………………………………………..………………………………….…..………………………..………………………

1. *Nội dung các phần thuyết minh và tính toán:*

…...………………………………………………………………………………………

…...………………………………………………………………………………………

…...………………………………………………………………………………………

…...………………………………………………………………………………………

…...………………………………………………………………………………………

1. *Các bản vẽ, đồ thị ( ghi rõ các loại và kích thước bản vẽ ):*

…...………………………………………………………………………………………

…...………………………………………………………………………………………

…...………………………………………………………………………………………

…...………………………………………………………………………………………

1. *Họ tên người hướng dẫn:* …………………………………..……………………
2. *Ngày giao nhiệm vụ đồ án:*  *……../……./201…..*
3. *Ngày hoàn thành đồ án: ……../……./201…..*

|  |  |
| --- | --- |
|  | *Đà Nẵng, ngày tháng năm 201* |
| **Trưởng Bộ môn** …………………….. | **Người hướng dẫn** |

**LỜI NÓI ĐẦU**

{Font: Time New Roman; thường; cỡ chữ: 13; dãn dòng: 1,3; căn lề: justified} ………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

………………………………………………………………………………..…………

Ghi chú: Sinh viên có thể trình bày “Lời cảm ơn” trong phần “Lời nói đầu”

**CAM ĐOAN**

{Font: Time New Roman; thường; cỡ chữ: 13; dãn dòng: 1,3; căn lề: justified}

{Lời cam đoan của sinh viên thực hiện đồ án tốt nghiệp cam đoan về liêm chính học thuật}

……………………………………………………………………..……………………

……………………………………………………………………..……………………

……………………………………………………………………..……………………

……………………………………………………………………..……………………

……………………………………………………………………..……………………

Sinh viên thực hiện

{Chữ ký, họ và tên sinh viên}

**MỤC LỤC**

*{Để 2 dòng trống tại đây}*

*{Font: Time New Roman; thường; cỡ chữ: 13; dãn dòng: 1,3; căn lề: justified}{In trên 2 mặt giấy từ trang này đến hết phần “PHỤ LỤC”}*

Tóm tắt

Nhiệm vụ đồ án

Lời nói đầu và cảm ơn i

Lời cam đoan liêm chính học thuật ii

Mục lục iii

Danh sách các bảng biểu, hình vẽ và sơ đồ v

Danh sách các cụm từ viết tắt vi

Trang

[Chương 1: TỔNG QUAN ĐỀ TÀI 3](#_Toc9605719)

[1.1. Giới thiệu đề tài 3](#_Toc9605720)

[1.2. Mục đích đề tài 3](#_Toc9605721)

[1.3. Yêu cầu đề tài 4](#_Toc9605722)

[Chương 2: CƠ SỞ LÝ THUYẾT 5](#_Toc9605723)

[2.1. Mô tả hệ thống 5](#_Toc9605724)

[2.1.1. Nhận biết ngữ cảnh 5](#_Toc9605725)

[2.2. Hệ gợi ý 6](#_Toc9605726)

[2.2.1. Định nghĩa 7](#_Toc9605727)

[2.2.2. Dự đoán trong hệ gợi ý 7](#_Toc9605728)

[2.3. Các phương pháp truyền thống 7](#_Toc9605729)

[2.3.1. Gợi ý dựa trên nội dung 9](#_Toc9605730)

[2.3.2. Gợi ý dựa trên cộng tác 10](#_Toc9605731)

[2.3.3. Gợi ý dựa trên cách tiếp cận kết hợp 11](#_Toc9605732)

[2.4. Các cách tiếp cận gợi ý hiện đại 12](#_Toc9605733)

[2.4.1. Tiếp cận nhận biết ngữ cảnh 12](#_Toc9605734)

[2.4.2. Tiếp cận dựa trên ngữ nghĩa 12](#_Toc9605735)

[Chương 3: THIẾT KẾ VÀ CÀI ĐẶT HỆ THỐNG 15](#_Toc9605736)

[3.1. Thiết kế hệ thống 15](#_Toc9605737)

[3.1.1. Mô tả các chức năng 15](#_Toc9605738)

[3.1.2. Các sơ đồ hệ thống 16](#_Toc9605739)

[3.2. Cài đặt 18](#_Toc9605740)

[3.2.1. Mở giao diện mặc định 18](#_Toc9605741)

[3.2.2. Thực thi thuật toán 19](#_Toc9605742)

[3.2.3. Xây dựng các thành phần chức năng 22](#_Toc9605743)

[Chương 4: KẾT QUẢ THỰC NGHIỆM 23](#_Toc9605744)

[4.1. Quy trình sinh dữ liệu 23](#_Toc9605745)

[4.2. Triển khai hệ thống 23](#_Toc9605746)

[4.3. Kết quả thực nghiệm 23](#_Toc9605747)

**DANH SÁCH CÁC BẢNG, HÌNH VẼ**

**Danh sách hình vẽ**

[Hình 3.1. Sơ đồ phân cấp chức năng của ứng dụng 16](#_Toc9607361)

[Hình 3.2. Sơ đồ luồng dữ liệu của ứng dụng 16](#_Toc9607362)

**Danh sách bảng**

[Bảng 3.1. Danh sách các trang ứng với đường dẫn của ứng dụng 17](#_Toc9026481)

***Ghi chú:***

* Mỗi bảng, hình vẽ/ sơ đồ phải được đánh số và có tên;
* Đánh số bảng và đánh số hình vẽ/ sơ đồ riêng. Quy luật đánh số như sau:
  + Chữ số thứ nhất chỉ tên chương;
  + Chữ số thứ hai chỉ thứ tự bảng biểu, sơ đồ, hình,…trong mỗi chương.

**DANH SÁCH CÁC KÝ HIỆU, CHỮ VIẾT TẮT**

KÝ HIỆU:

…….……..........................................................................................................................

…….……..........................................................................................................................

…….……..........................................................................................................................

…….……..........................................................................................................................

…….……..........................................................................................................................

…….……..........................................................................................................................

CHỮ VIẾT TẮT:

RS: Recommender Systems

CARS: Context Aware Recommender Systems

CAMF: Context Aware Matrix Factorization

CSLIM: Contextual Sparse Linear Method

TF: Tensor Factorization

MF: Matrix Factorization

*Ghi chú:*

* Ký hiệu: mỗi mục ký hiệu gồm ký hiệu và phần tên gọi, diễn giải ký hiệu.
* Cụm từ viết viết tắt là các chữ cái và các ký hiệu thay chữ được viết liền nhau, để thay cho một cụm từ có nghĩa, thường được lặp nhiều lần trong đồ án.

**MỞ ĐẦU**

Hiện nay, việc mua bán online đang diễn ra vô cùng mạnh mẽ. Mỗi website có hàng ngàn, chục ngàn đến trăm ngàn sản phẩm với nhiều loại và công dụng khác nhau. Tương tự như vậy thì các website hướng dẫn du lịch cũng như vậy, khách hàng của họ chỉ ghé qua trong thời gian ngắn nhưng họ lại muốn thấy ngay những gì họ cần, nếu không họ sẽ cảm thấy thật khó khăn trong việc tìm cái mình cần ở một website. Từ đó, việc đưa ra gợi ý một cách vô cùng hợp lý là điều mà website nào cũng cần. Ngay khi khách hàng truy cập vào website và thực hiện một vài thao tác đầu tiên, việc đưa ra ngay gợi ý cho khách hàng mặc hàng tiếp theo là vô cùng cần thiết. Hiểu được khách hàng là một điều không phải ai cũng làm được nhất là đối với lập trình viên nhưng khách hàng lại chính là người có thể hiểu rõ nhau nhất. Chúng ta hoàn toàn có thể tìm điểm chung giữa họ và dự đoán cho những người có cùng điểm chung là họ cũng sẽ thích những mặc hàng mà những người có chung sở thích với họ cũng thích. Từ đó mà việc có một hệ gợi ý là một điều vô cùng quan trọng.

Với dữ liệu lịch sử khách hàng và các yếu tố khách quan, hệ gợi ý sẽ đưa ra các mặc hàng tiếp theo mà dự đoán rằng khách hàng sẽ quan tâm đến nó. Và đưa các gợi ý đó lên đầu trang để thu hút khác hàng về website của mình. Mục tiêu của đề tài là giúp cho bên nhà cung cấp dịch vụ có thể tìm ra hệ gợi ý phù hợp nhất với dữ liệu vào (lịch sử khách hàng) của mình và từ đó đưa ra các gợi ý phù hợp cho các khách hàng trong tương lai. Hệ gợi ý phù hợp bao gồm thuật toán phù hợp, thông số cài đặt phù hợp, thời gian chạy phù hợp và kết quả ra với xác suất cao nhất có thể của hệ.

Hệ thống nhắm đến mọi website có hình thức dịch vụ mua bán và giới thiệu sản phẩm cho người dùng. Để người dùng nhận được gợi ý hợp lý hay chính nhà cung cấp có thể tìm được người dùng thích hợp để giới thiệu. Và cũng như mọi hệ gợi ý khác, gợi ý không chỉ quan trọng mà còn cần phải đúng và đủ cho từng người dùng vì vậy mà tìm ra được hệ phù hợp là việc không phải dễ và cũng không dễ để áp dụng. Từ đó mà việc xây dựng một ứng dụng giúp bên thứ ba dễ dàng giải quyết bài toán đau đầu này là điều cần thiết.

{Font: Time New Roman; thường; cỡ chữ: 13; dãn dòng: 1,3; căn lề: justified}

{Cấu trúc của đồ án tốt nghiệp}

# TỔNG QUAN ĐỀ TÀI

* 1. Giới thiệu đề tài

Hiện nay, việc lựa chọn nơi du lịch thích hợp theo sở thích, giá tiền và thời tiết (có thể xem như là thời điểm) thường sẽ tốn thời gian, ngoài ra có một số người không nhận ra rằng có 1 số địa điểm mà họ cũng có thể thích (địa điểm tương tự) và nhờ các phương pháp gợi ý mà các website có thể dễ dàng phán đoán nơi nào là lựa chọn tốt nhất cho khách hàng và thêm nữa là những nơi khách hàng có thể đi cùng lúc với nơi mình chọn. Nhưng không phải website nào cũng có hệ gợi ý và lựa chọn hệ gợi ý tốt nhất, có thể vì họ không có sự lựa chọn và cũng không dễ dàng để lựa chọn hệ tốt nhất cho mình. Vì vậy, đề tài mong giúp đỡ các website mua bán và quảng cáo có thể dễ dàng áp dụng và đưa ra các gợi ý thích hợp cho người dùng.

* 1. Mục đích đề tài

Với mục đích ban đầu là giúp cho người dùng (những người cần tìm ra hệ gợi ý hợp lí với họ) dễ dàng đánh giá các thuật toán gợi ý và tìm ra thuật toán tốt nhất, ứng dụng hệ gợi ý hướng ngữ cảnh người dùng ra đời với hi vọng sẽ giúp người dùng có thể thử nghiệm từng thuật toán, đánh giá so sánh và sau cùng là tạo ra được gợi ý cho hệ thống của mình.

* 1. Yêu cầu đề tài

Với dữ liệu sẵn có từ người, phần mềm sẽ tạo ra 1 giao diện thao tác để người dùng có cơ hội thử và đánh giá từng phương pháp gợi ý trong hệ gợi ý và đặc biệt có thể thử các phương pháp gợi ý mới nhất trong đó có hệ gợi ý theo hướng ngữ cảnh. Sau cùng, người dùng có thể tùy chỉnh thông số phù hợp để đạt được kết quả gợi ý tốt nhất và sau đó trích ra gợi ý dựa trên dữ liệu đầu vào thành gợi ý cho hệ thống của mình.

Yêu cầu của hệ thống:

* Đánh giá của người dùng về một sản phẩm nào đó (một chuyến đi) và các ngữ cảnh xunh quanh và từ chính người dùng.
* Thuật toán tìm ra gợi ý và các thông số cài đặt.

Kết quả của hệ thống:

* Kết quả độ chính xác của thuật toán xét trên nhiều thuật toán đánh giá khác nhau.
* Một loạt danh sách về các gợi ý mà hệ thống đưa ra cho từng người dùng thông qua bộ dữ liệu ban đầu.

# CƠ SỞ LÝ THUYẾT

* 1. Mô tả hệ thống
     1. Nhận biết ngữ cảnh
        1. Ngữ cảnh

Ngữ cảnh là "bất kỳ thông tin nào mà có thể sử dụng được để đặc tả một tình huống của một thực thể. Một thực thể là một người, một nơi hay một số đối tượng được xem là có liên quan đến tương tác giữa người dùng và ứng dụng, bao gồm cả chính người dùng và ứng dụng đó". Đồng thời, ông cũng cung cấp định nghĩa sau cho các hệ thống tính toán nhận biết ngữ cảnh: "Một hệ thống sử dụng ngữ cảnh để cung cấp các thông tin liên quan hoặc các dịch vụ cho người dùng trong đó mối liên quan phụ thuộc vào nhiệm vụ của người dùng".

* + - 1. Các thuộc tính của ngữ cảnh

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Kiểu** | **Nguồn** | **Tính lâu dài** | **Chất lượng** | **Nguyên nhân lỗi** |
| Cảm biến | Cảm biến vật lý và logic | Ngắn | Tùy vào chất lượng hệ thống | Lỗi thiết bị cảm biến, mất mạng, ... |
| Đặc tả | Do người dùng đặc tả trực tiếp hoặc gián tiếp | Dài | Không đảm bảo tính chính xác | Do cảm xúc, khả năng mô tả, lỗi ghi chép, ... |

Bảng 1: Các thuộc tính của các nguồn ngữ cảnh

* + - 1. Phân loại ngữ cảnh

|  |  |
| --- | --- |
| **Ngữ cảnh** | **Tham số ngữ cảnh** |
| Ngữ cảnh tĩnh của người dùng | Tên, tuổi, thói quen, sở thích, ... |
| Ngữ cảnh động của người dùng | Hành động, cảm xúc, trạng thái, ... |
| Ngữ cảnh môi trường | Thời tiết, thời điểm, địa điểm, ... |

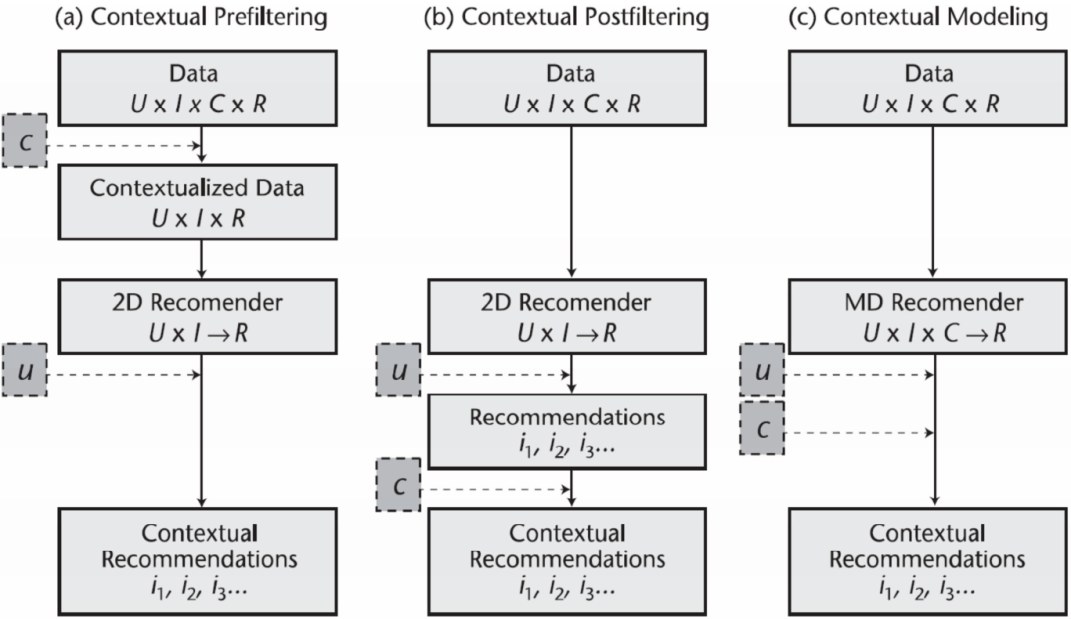
Bảng 2: Phân loại các chiều của ngữ cảnh

* + 1. Phương pháp

Trong nhệ thống gợi ý nhận biết ngữ cảnh, ngữ cảnh luôn luôn được xem là “bất cứ thông tin nào có thể được sử dụng để đặc trưng hóa tình hình của một thực thể”, bao gồm *thời gian* và *người đồng hành* có thể là hai điều có ảnh hưởng đến ngữ cảnh, bởi vì sở thích của người dùng có thể thay đổi khi hai ngữ cảnh kia thay đổi.

Giới thiệu về phân loại hai phần của thông tin theo ngữ cảnh. Phân loại của nó dựa trên hai sự cân nhắc: những điều mà hệ gợi ý nhận biết về một yếu tố ngữ cảnh (contextual factor) và cách mà yếu tố ngữ cảnh đó thay đổi theo thời gian. Hiểu biết của hệ thống (system knowledge) có thể được chia thành: *quan sát được hoàn toàn* (fully observable), *quan sát được một phần* (partially observable) và *không quan sát được* (unobservable). Khía cạnh thời gian của một yếu tố có thể chia thành *tĩnh* (static) và *động* (dynamic).

Một vài thuật toán gợi ý nhận biết ngữ cảnh đã được phát triền trong vài thập kỉ trước. Điển hình là, ngữ cảnh có thể áp dụng trong ba chiến thuật cơ bản sử dụng gợi ý: *lọc trước* (pre-filtering), *lọc sau* (post-filtering) và *mô hình ngữ cảnh* (contextual modeling). Kỹ thuật pre-filtering, như là cách tiếp cận phân tách nhận biết ngữ cảnh (Context-Aware Splitting Approaches – CASA), áp dụng ngữ cảnh làm bộ lọc trước để lọc ra các hồ sơ xếp hạng không liên quan. Post-filtering, mặc khác, áp dụng ngữ cảnh như bộ lọc cho hệ gợi ý sau quá trình ra gợi ý. Trong contextual modeling, các mô hình dự đoán được học cách sử dụng toàn bộ ngữ cảnh. Hầu hết chúng vẫn còn hoạt động, như là phân rã ma trận nhận diện ngữ cảnh (CAMF), phân rã nhân tử (TF) và phương pháp tuyến tính ngữ cảnh thưa thớt (CSLIM).



Hình : Chiến thuật sử dụng ngữ cảnh

* 1. Hệ gợi ý
     1. Định nghĩa

Hệ gợi ý (Recommender system - RS) là một phần cách thuận tiện để giảm thiểu lượng thông tin thừa trong nhiều lĩnh vực ứng dụng bằng cách điều chỉnh các khuyến nghị cho người dùng theo sở thích cá nhân. Ngoài ra còn có một số loại hệ thống gợi ý xuất hiện trong những năm gần đây, chẳng hạn như hệ thống gợi ý hướng ngữ cảnh (Context Aware Recommender Systems – CARS) và gợi ý bối cảnh (Context Recommender). CARS nổi lên vượt ra ngoài sở thích người dùng và cũng tính đến tình huống bối cảnh của người dùng trong việc tạo ra đề xuất. CARS đã chứng minh hiệu quả trong một số lĩnh vực, chẳng hạn như du lịch, âm nhạc, phim và nhà hàng. Lấy

* + - 1. Dự đoán trong hệ gợi ý truyền thống

Công thức tiêu chuẩn để giải quyết vấn đề gợi ý bắt đầu với một ma trận xếp hạng hai chiều, được tạo nên từ người dùng và sản phẩm:

R: Users × Items → Ratings.

* + - 1. Dự đoán trong hệ gợi ý có ngữ cảnh

Chìa khóa nằm bên trong của các hệ thống gợi ý nhận biết ngữ cảnh là các ưa thích của người dùng cho các sản phẩm có thể bị bối cảnh tác dụng một phần lên các sản phẩm đó. Việc kết hợp bối cảnh yêu cầu cần ước tính sở thích của người dùng bằng cách sử dụng hàm xếp hạng đa chiều (a multidimensional rating function):

R: Users × Items × Contexts → Ratings

* 1. Các phương pháp truyền thống

Gọi U(users) là tập người dùng, I(items) là tập các sản phẩm có thể được gợi ý. Để dự đoán xếp hạng (hay tính tiện ích) của sản phẩm i đối với người dùng u thì người ta đưa ra hàm xếp hạng (rating) r:

U x I → R

Trong đó: R(recommendations) là tập các giá trị xếp hạng được thứ tự toàn phần (ví dụ: số nguyên dương hoặc số thực trong tập xác định). Mô hình này còn được gọi với tên gọi là mô hình dự đoán 2 chiều (two-dimensional recommendation framework).

Hình 1: Các thành phần cơ bản của tiến trình gợi ý truyền thống

Với mỗi người dùng u ∈ U, chúng ta có thể chọn được sản phẩm i ∈ I sao cho hàm xếp hạng của người dùng u đối với item i là lớn nhất.

∀𝑢 ∈ 𝑈, 𝑖 ′ 𝑢 = arg max 𝑟̂(𝑢, 𝑖)

𝑖∈𝐼

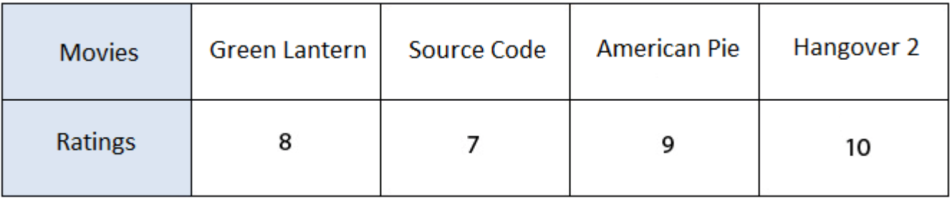
Có rất nhiều phương pháp tiếp cận trong hệ gợi ý, tuy nhiên có thể chia thành 3 nhóm kỹ thuật chính như sau:

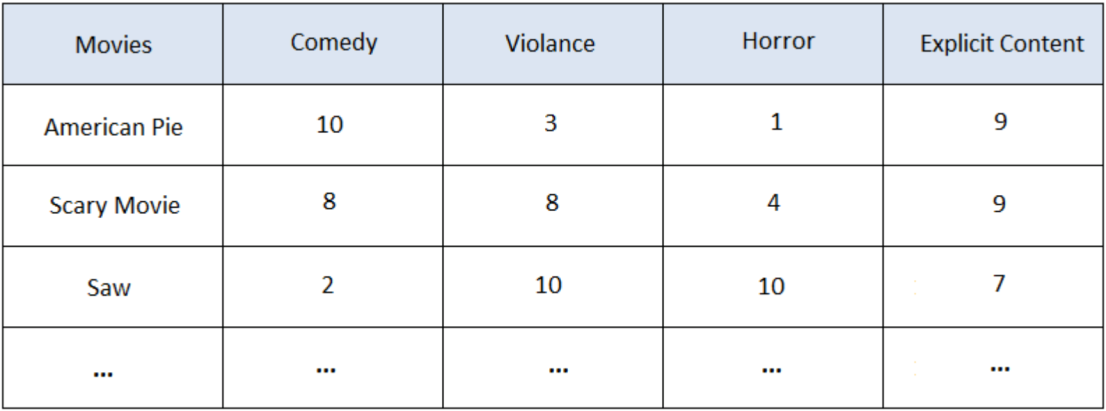
* Gợi ý dựa trên nội dung (content-based filtering)
* Gợi ý dựa trên cộng tác (collaborative filtering - CF)
* Gợi ý dựa trên cách tiếp cận kết hợp (hybrid approach)
  + 1. Gợi ý dựa trên nội dung

Hệ thống gợi ý dựa trên nội dung làm việc với hồ sơ của người dùng được tạo ngay từ lúc ban đầu. Một hồ sơ có thông tinh về một người dùng và “khẩu vị”(taste) của người đó. Khẩu vị được dựa trên cách người dùng đánh giá các sản phẩm. Thông thường, khi tạo một hồ sơ, hệ thống gợi ý tạo một khảo sát, để lấy thông tin khởi đầu về một người dùng để tránh những vấn đề với người dùng mới.

Trong quá trình đưa ra gợi ý, hệ thống so sánh những sản phẩm đã được đánh giá tích cực bởi người dùng với những sản phẩm người đó không đánh giá và tìm kiếm điểm tương đồng. Những sản phẩm mà gần như tương đồng với cái được đánh giá tích cực, sẽ được gợi ý cho người dùng.

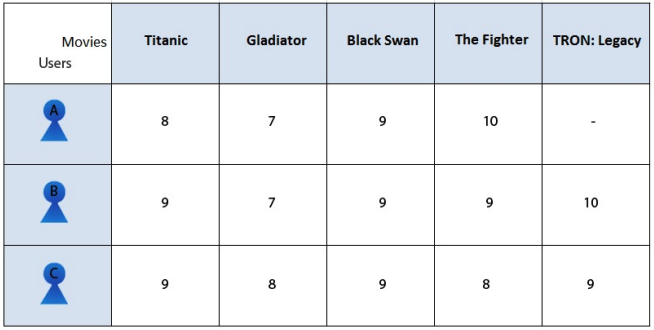
Hình 3.1 cho thấy một ví dụ cho một hồ sơ của người dùng với những bộ phim mà người đó đã xem và người đó đã đánh giá. Hình 3.2 cho thấy danh sách những bộ phim và tính chất của chúng. Một hệ thống gợi ý dựa trên nội dung sẽ tìm ra những bộ phim từ danh sách (hình 3.2) và tìm kiếm những điểm tương đồng. Những bộ phim tương đồng sẽ được gợi ý cho người dùng. Trong ví dụ này chúng ta có thể thấy rằng bộ phim “Scary Movie” tương tự với bộ phim “American Pie” mà người xe đã đánh giá tích cực vậy nên nó sẽ được gợi ý cho người dùng này.

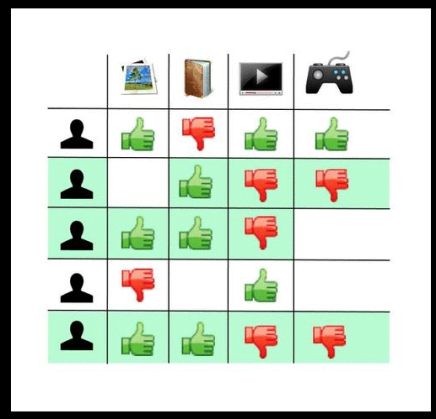
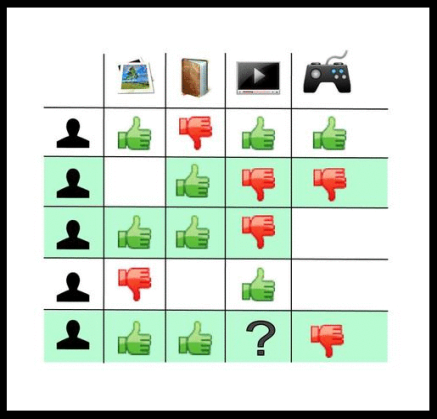
  
Hình 3.1: Lịch sử đánh giá của người dùng

  
Hình 3.2: Danh sách các bộ phim

* + 1. Gợi ý dựa trên cộng tác

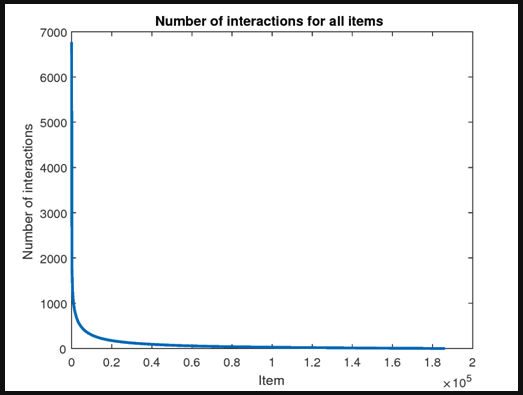
Gợi ý cộng tác trở thành một trong hầu hết các công nghệ của hệ thống gợi ý được nghiên cứu từ khi cách tiếp cận này được đề cập đến và mô tả bởi Paul Resnick và Hal Varian vào năm 1007. Ý tưởng về lọc cộng tác là tìm kiếm người dùng trong một cộng đồng chia sẻ đánh giá. Nếu hai người dùng có cùng hoặc có cùng hầu hết đánh giá về những sản phẩm chung, thì họ có cùng “khẩu vị”. Giống như những người dùng tạo nên một nhóm hoặc gọi là những người hàng xóm. Một người dùng nhận được gợi ý đến những sản phẩm mà họ chưa đánh giá trước đó, nhưng đã được đánh giá tích cực bởi hàng xóm của họ. Ảnh 3.3 cho thấy rằng cả 3 người dùng đều đánh giá các bộ phim tích cực với cùng số điểm. Có nghĩa là họ có cùng khẩu vị và được nhóm vào 1 nhóm hàng xóm. Người dùng A chưa đánh giá bộ phim “TRON: Legacy”, mà có lẽ người đó chưa xem. Vì bộ phim đã được đánh giá cao bởi những hàng xóm khác, anh ấy sẽ được gợi ý bộ phim đó. Trái ngược với với hệ gợi ý đơn giản nơi mà gợi ý dựa trên sản phẩm đã được đánh giá nhiều và sản phẩm phổ biến nhất, gợi ý cộng tác quan tâm đến khẩu vị của người dùng. Khẩu vị được đánh giá là không đổi hoặc đổi rất ít.

  
Hình 3.3: Ví dụ hệ gợi ý hợp tác

  
Hình 3.4: Ví dụ về hệ gợi ý hợp tác trên wiki

* + 1. Gợi ý dựa trên cách tiếp cận kết hợp

Để kết có kết quả tốt hơn, vài hệ thống gợi ý kết hợp các công nghệ khác nhau của cách tiếp cận cộng tác và cách tiếp cận dựa trên nội dung. Sử dụng cách tiếp cận kết hợp, ta có thể bỏ qua giới hạn và các vấn đề của một hệ thống yếu, như là “cold-start problem”.

  
Hình 3.5: Cold-start problem

Sự kết hợp của các phương pháp có thể thực hiện theo các cách khác nhau:

* Thực hiện đầy đủ các thuật toán và gộp kết quả.
* Sử dụng một số quy tắc lọc dựa trên nội dung trong phương pháp hợp tác
* Sử dụng một số quy tắc lọc cộng tác theo cách tiếp cận lọc nội dung
* Tạo một hệ thống đề xuất thống nhất, kết hợp cả hai cách tiếp cận
  1. Các cách tiếp cận gợi ý hiện đại

Trong vài thập kỉ đã qua, một số thuật toán gợi ý nhận diện ngữ cảnh đã được phát triển, như là:

* Mô hình bối cảnh khác biệt (Differential Context Modeling): kết hợp ngữ cảnh như là những bộ lọc trong lọc cộng tác (Collaborative Filtering).
* Phân rã ma trận nhận diện ngữ cảnh (Context-Aware Matrix Factorization - CAMF): sử dụng ngữ cảnh trong phân rã ma trận.
* Phân rã nhân tử (Tensor Factorization - TF): trực tiếp xem xét ngữ cảnh là chiều cá nhân trong không gian xếp hạng đa chiều.
  + 1. Tiếp cận nhận biết ngữ cảnh

Ngữ cảnh là thông tin về môi trường của một người dùng và tình huống cụ thể của người đó. Những chi tiết như vậy có thể đáng kể hơn nhiều luật trong gợi ý hơn việc đánh giá các sản phẩm, vì việc đánh giá, một mình nó không có thông tin chi tiết về hoàn cảnh bên dưới mà nó được nhận bởi người dùng. Vài gợi ý có thể thích hợp hơn với người dùng trong buổi chiều và không hợp với sở thích của người đó vào buổi sáng, và người dùng muốn làm một việc khi trời lạnh nhưng lại không làm khi trời nóng. Hệ gợi ý chú ý và sử dụng thông tin đó trong việc đưa ra các khuyến nghị được gọi là hệ thống gợi ý nhận biết ngữ cảnh.

* + 1. Tiếp cận dựa trên ngữ nghĩa

Hầu hết thông tin của sản phẩm, người dùng trong hệ gợi ý và phần còn lại của trang web được thể hiện ra trong trang web ở dạng văn bản. Sử dụng chủ đề và từ khóa mà không có bất kỳ ý nghĩa ngữ nghĩa nào sẽ không cải thiện tính chính xác của các đề xuất trong mọi trường hợp, vì một số từ khóa có thể là từ đồng âm. Đó là lý do tại sao việc hiểu và cấu trúc cho văn bản là một khuyến nghị phần rất quan trọng. Phương pháp khai thác văn bản truyền thống dựa trên phân tích từ vựng và cú pháp cho thấy các mô tả có thể hiểu được bởi người dùng chứ không phải máy tính hoặc hệ thống đề xuất. Đó là một lý do tạo ra các kỹ thuật xác định ý nghĩa văn bản mới dựa trên phân tích ngữ nghĩa. Các hệ thống đề xuất với các kỹ thuật như vậy được gọi là các hệ thống đề xuất ngữ nghĩa.

* 1. Các thuật toán
     1. Phân rã ma trận – Matrix factorization

Phân rã ma trận là một trong các thuật toán gợi ý hiệu quả nhất trong hệ thống gợi ý truyền thống. Đơn giản, cả người dùng và sản phẩm đều được biểu diễn dưới dạng vector. Ví dụ, được sử dụng để biểu thị vector người dùng, như là một vector sản phẩm. Các giá trị trong những vector đó có thể được biểu diễn bởi phương trình 1.

(1)

Đặc biệt hơn, trọng sô của được sử dụng để biểu diễn số lượng người dùng thích các yếu tố tìm ẩn đó, và trọng số thể hiện cách mà sản phẩm này chứa các yếu tố tìm ẩn đó. Do đó, dấu chấm tích của hai vector này được sử dụng để biểu lộ mức độ người dùng thích sản phẩm này?

(2)

Bên cạnh đó, người dùng và thiên vị đánh giá sản phẩm cũng được thêm vào, như đã biểu diễn ở công thức số 2?, nơi mà biểu lộ đánh giá trung bình toàn bộ trong bộ dữ liệu, và biểu diễn người dùng và thiên vị cho sản phẩm tương ứng.

* + 1. Context-aware Matrix Factorization

# THIẾT KẾ VÀ CÀI ĐẶT HỆ THỐNG

* 1. Thiết kế hệ thống
     1. Mô tả các chức năng
        1. Tùy chỉnh các đường dẫn thư mục

Hệ thống yêu cầu 1 tập tin setting.conf để có thể lưu các đường dẫn và cài đặt đầu vào cho thuật toán. Ngoài ra có thể tùy chỉnh nơi lưu tệp dữ liệu của người dùng để chạy thuật toán.

* + - 1. Chạy các thuật toán

Người dùng có thể chạy cả thuật toán không sử dụng ngữ cảnh cũng như thuật toán có sử dụng ngữ cảnh.

* + - 1. Điều chỉnh thông số cho hệ

Với từ hệ gợi ý sẽ có các thông số đầu vào, người dùng có thể tùy chỉnh để tối ưu kết quả đạt được của thuật toán với mội bộ dữ liệu khác nhau.

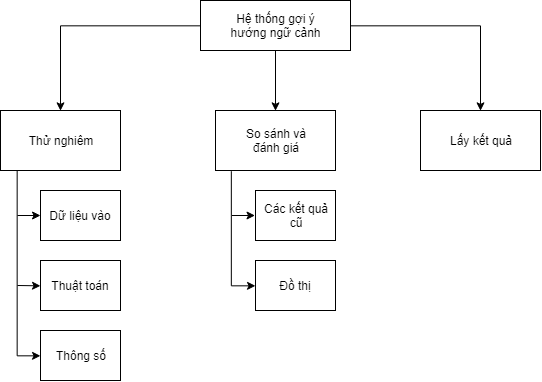
* + - 1. So sánh giữa kết quả của các thuật toán

Phần mềm sẽ lưu lại kết quả tốt nhất của thuật toán và từ đó cho người dùng cái nhìn tổng quan nhằm lựa chọn thuật toán tốt nhất cho hệ thống.

* + - 1. Lấy các kết quả gợi ý

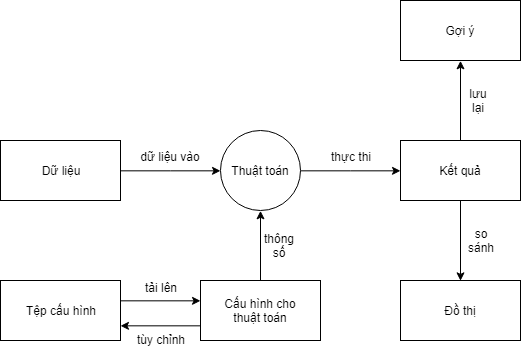
Với mỗi thuật toán sẽ sinh ra một loạt các kết quả gợi ý mặt hàng thích hợp cho người dùng được lưu vào các tệp để người dùng có thể sử dụng.

* + 1. Các sơ đồ hệ thống
       1. Sơ đồ phân cấp chức năng



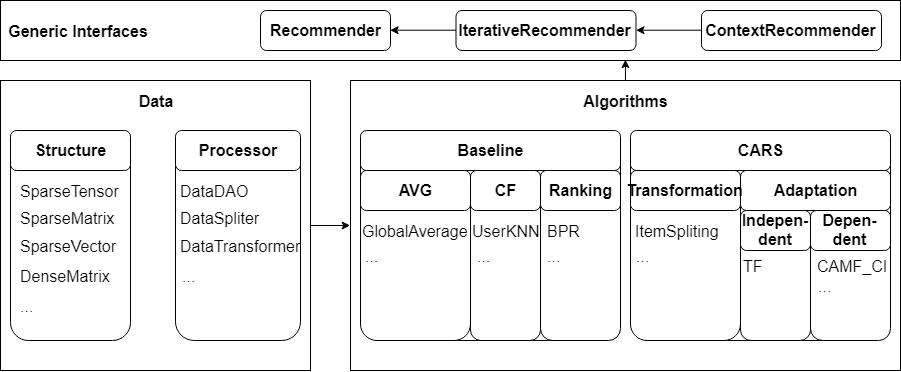
Hình 3.1. Sơ đồ phân cấp chức năng của ứng dụng

* + - 1. Sơ đồ luồng dữ liệu



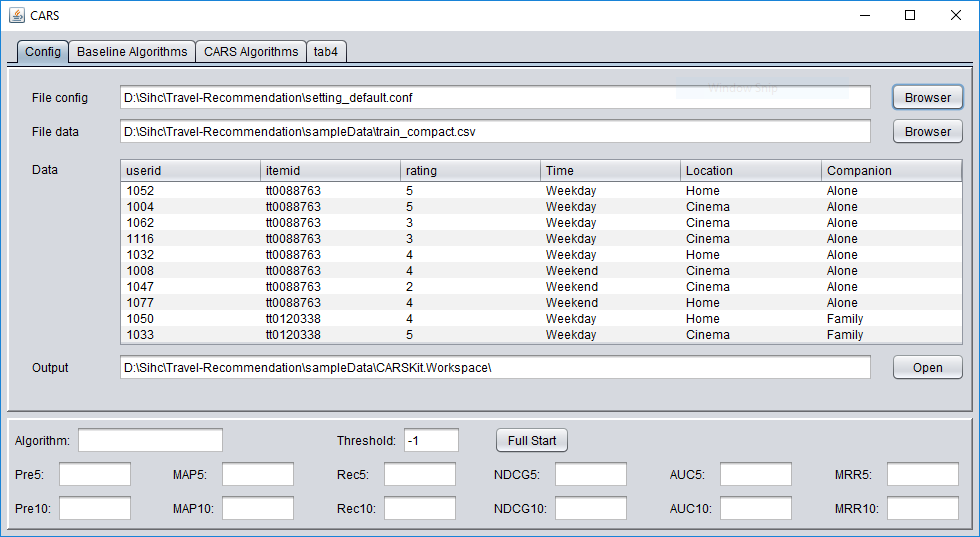
Hình 3.2. Sơ đồ luồng dữ liệu của ứng dụng

* + - 1. Sơ đồ thuật toán



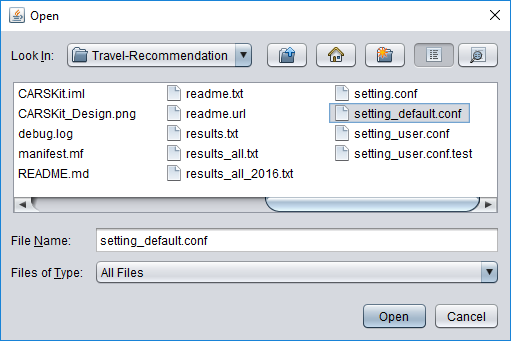
Hình 3.3: Sơ đồ dữ liệu thuật toán

* 1. Cài đặt
     1. Mở giao diện mặc định



Hình 3.4: Giao diện khởi động của chương trình

Ở giao diện khởi động, người dùng sẽ chọn nơi chứa tệp setting.conf của mình để hệ thống nhận. Thông thường, hệ thống sẽ tự mở tập setting\_default.conf làm mặc định.

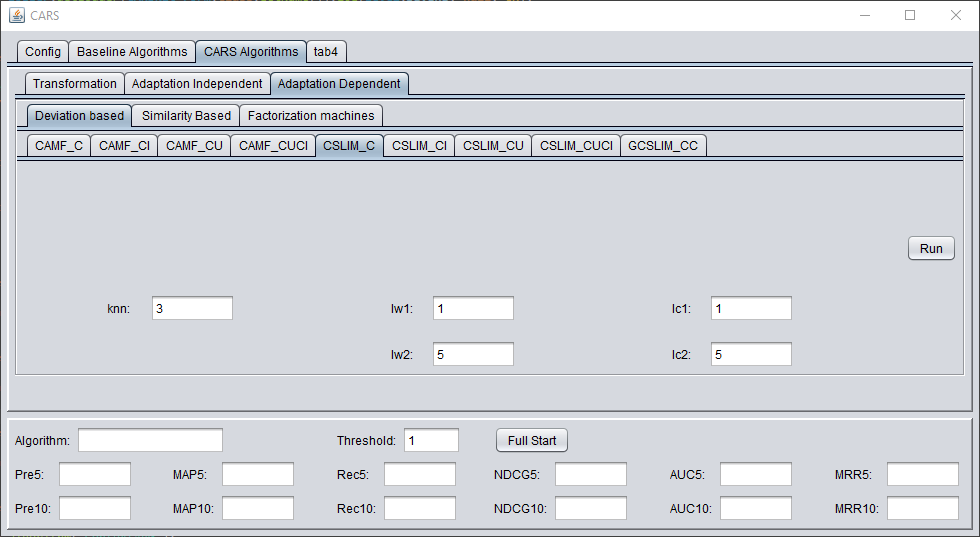


Hình 3.5: Browser để tìm tệp setting.conf phù hợp

Sau đó người dùng sẽ chọn tệp dữ liệu đầu vào tại File data. Sau khi chọn tệp, dữ liệu sẽ được biểu diễn ở bên dưới, người dùng có thể xem lại.

Cuối cùng là Output là đường dẫn nơi kết quả dự đoán của từng thuật toán được đưa ra, sau khi xác định thuật toán tốt nhất, người dùng sẽ ở thư mục Output và lấy kết quả.

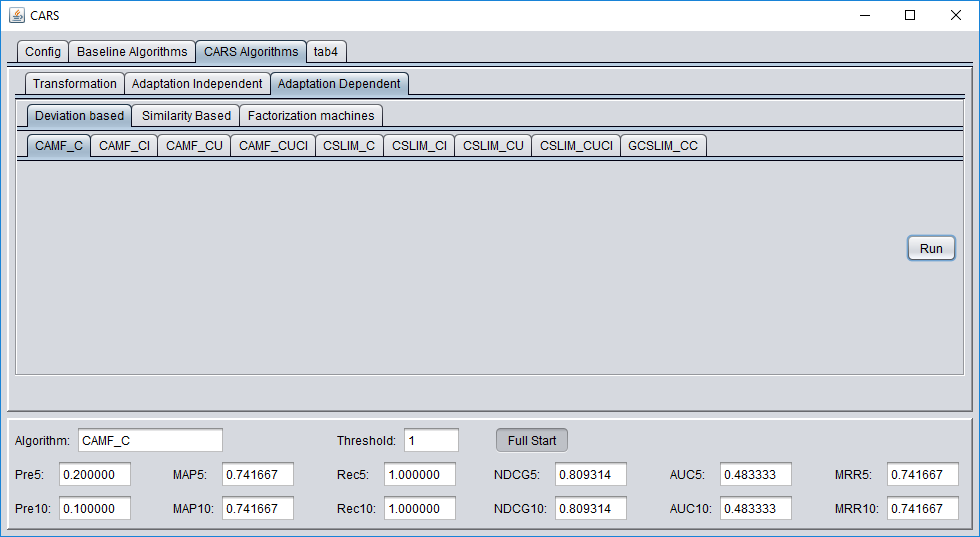
* + 1. Thực thi thuật toán
       1. Thiết lập thông số



Hình 3.6: Thuật toán CSLIM\_C có 5 thông số đầu vào

Có một số thuật toán sẽ yêu cầu thông số đầu vào. Thông thường sẽ có các thông số mặc định nhưng người dùng có thể tinh chỉnh để tối ưu them kết quả đạt được.

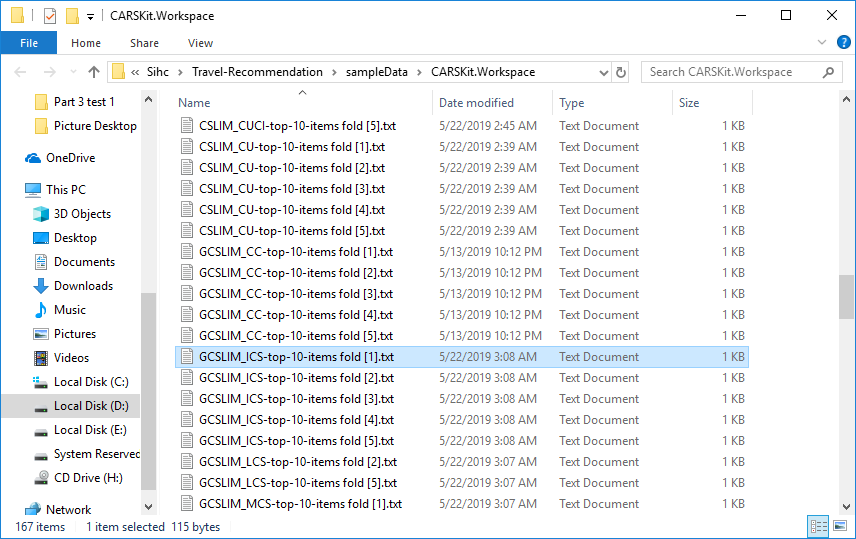
* + - 1. Chạy thuật toán



Hình 3.7: Kết quả sau khi chạy xong 1 thuật toán

Sau khi thực hiện xong một thuật toán, sẽ có kết quả hiện ở phần kết quả. Kết quả là một loạt các đánh giá bằng các thuật toán đánh giá xếp hạng khác nhau nhằm tạo cho người dùng một cái nhìn tổng quát về chất lượng thuật toán.

* + - 1. Xem lịch sử
      2. Mở thư mục chứa gợi ý



*Hình 3.8: Thư mục chứa gợi ý sau khi chạy xong thuật toán*

Mỗi thuật toán sau khi chạy xong sẽ sinh ra 1 số lượng file nhất định (tùy chỉnh trong tệp config) dựa trên số lần mình chạy lại thuật toán. Sau khi lựa chọn được thuật toán phù hợp nhất cho bộ dữ liệu thì chúng ta có thể lấy kết quả đó làm gợi ý cho khách hang.

* + 1. Xây dựng các thành phần chức năng

# KẾT QUẢ THỰC NGHIỆM

* 1. Quy trình sinh dữ liệu
  2. Triển khai hệ thống
  3. Kết quả thực nghiệm