TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

**BÀI TẬP LỚN MÔN KHAI PHÁ WEB**

**TÀI LIỆU KẾ HOẠCH THỰC HIỆN**

**Đề Tài: Xây Dựng Hệ Gợi Ý Phim Sử Dụng**

**Item-Based Collaborative Filtering**

|  |  |
| --- | --- |
|  |  |
|  |  |

*Hà Nội, 03/2020*

# Đặt vấn đề , lý do chọn đề tài

## Hệ gợi ý là gì?

Hệ gợi là một phần của một hệ thống thông tin . Mục đích của nó là hỗ trợ người dùng tìm kiếm được đúng thông tin cần thiết, dự đoán sở thích hay xếp hạng mà người dùng có thể dành cho một mục thông tin (item) nào đó mà họ chưa gặp trong quá khứ.

## Lý do chọn đề tài

Dưới đây là một số ví dụ về công dụng của hệ gợi ý trong các hệ thống cụ thể :

• Youtube tự động chuyển các clip liên quan đến clip bạn đang xem. Youtube cũng tự gợi ý những clip mà có thể bạn sẽ thích.

• Khi bạn mua một món hàng trên Amazon, hệ thống sẽ tự động gợi ý “Frequently bought together”, hoặc nó biết bạn có thể thích món hàng nào dựa trên lịch sử mua hàng của bạn.

• Facebook hiển thị quảng cáo những sản phẩm có liên quan đến từ khoá bạn vừa tìm kiếm.

• Facebook gợi ý kết bạn.

• Netflix tự động gợi ý phim cho người dùng.

Việc sử dụng hệ gợi ý giúp cho hiệu quả của việc Marketing tăng lên một cách đáng kể . Người dùng có được sự thuận tiện nhanh chóng phù hợp với nhu cầu của mình , người cung cấp sản phẩm tăng được doanh thu.

## Cách tiếp cận

Có 2 cách tiếp cận chính để xây dựng một hệ gợi ý :

1. Content-based systems: đánh giá đặc tính của items được recommended. Ví dụ: một user xem rất nhiều các bộ phim về cảnh sát hình sự, vậy thì gơi ý một bộ phim trong cơ sở dữ liệu có chung đặc tính hình sự tới user này, ví dụ phim Người phán xử.

2. Collaborative filtering : hệ thống gợi ý items dựa trên sự tương quan (similarity) giữa các users và/hoặc items . Đối với phương pháp này, chúng ta có thể chi ra thành 2 cách tiếp cận nhỏ hơn đó là. user-base và item-base ( Ở đây nhóm chúng em chọn hướng tiếp cận là item-base )

## Đề xuất đề tài

Với một hệ thống có rất nhiều nội dung , người dùng có thể tiếp cận tới nội dung họ cần thông qua việc tìm kiếm , nhưng đôi khi họ không biết tìm kiếm gì vì vậy chúng ta cần áp dụng hệ gợi ý vào những hệ thống như vậy .

Ví dụ : Trong hệ thống web xem phim chúng ta có thể gợi ý cho người dùng những bộ phim có nôi dung tương tự với những phim người dùng đã xem và đánh giá cao .

Với những ưu điểm của hệ gợi ý đối với hệ thống cung cấp sản phẩm cho người dùng nhóm chúng em xin được chọn đề tài Hệ gợi ý phim với Item-Based Collaborative Filtering để thực hiện trong học phần Khai Phá Web.

## Yêu cầu đạt được

• Hiểu được hệ gợi ý là gì ?

• Cách hoạt động của hệ gợi ý , và cách áp dụng vào 1 hệ thống cụ thể.

• Xây dựng được 1 hệ gợi ý hoàn chỉnh .

# Mô tả đề tài

Hệ thống có chức năng gợi ý những bộ phim phù hợp với 1 user cụ thể . Ví dụ người dùng thích phim “ Kẻ huỷ diệt ” hệ thống sẽ gợi ý cho người này phim có nội dung tương tự với phim này .

Chúng em sử dụng:

* Bài báo: **“Item-Based Collaborative Filtering Recommendation Algorithms**” của Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl – 2001 làm tài liệu tham khảo chính.
* Tập dữ liệu: MovieLens 100k dataset là bộ dữ liệu bao gốm đánh giá của 943 người dùng cho 1682 bộ phim , mỗi người dùng đánh giá ít nhất 20 bộ phim.
* Phương pháp sử dụng: Item-Based Collaborative Filtering và thuật toán K-NN .
* Các bước triển khai :

• Biểu diễn mỗi item bằng một vector người dùng.

• Tính độ tương tự giữa các item

• Đối với người dùng A, tìm các item tương tự với các item A đã tương tác.

• Recommend item cho A từ các item nói trên, bằng các tiêu chí như trọng số cao, nhiều người tương tác …

# Giới thiệu công nghệ sử dụng và phân công công việc

## Công nghệ sử dụng

Ngôn ngữ cài đặt: Python

Các thư viện sử dụng : Pandas, Sklearn

## Phân công công việc và kế hoạch làm việc

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |
|  |  |  |
|  |  |
|  |  |  |
|  |  |  |