



# RECOMMENDATION SYSTEM

GV Hướng Dẫn: Ts. Lê Anh Cường

Nguyễn Sang Sinh – 51900771

Nguyễn Văn Dũng - 51900046

# CONTENT



## **I. Nội dung lý thuyết**

- Giới thiệu bài toán Recommendation System
- Mục tiêu bài toán Recommendation System

## **II. Phương pháp giải quyết bài toán**

- Tiếp cận lọc theo nội dung
- Tiếp cận theo hướng bộ lọc cộng tác
- Tiếp cận theo phương pháp lai

## **III. Đề xuất phương pháp giải quyết**

## Giới thiệu bài toán

---

Hệ thống Internet đang ngày càng được nâng cao về chất lượng, cùng với đó số lượng người sử dụng Internet cũng không ngừng gia tăng.

Thị trường thương mại điện tử cũng đang trên đà phát triển nhanh hơn bao giờ hết, với lượng dữ liệu mà người dùng cung cấp tính theo đơn vị mỗi giây đã và đang gây ra một thách thức tiềm ẩn về tình trạng quá tải thông tin, gây cản trở người dùng tiếp cận kịp thời các mục tiêu mà họ quan tâm trên Internet

## Giới thiệu bài toán

---

Hệ thống gợi ý (Recommendation System) là một hệ thống lọc thông tin giải quyết vấn đề quá tải thông tin bằng cách lọc lấy phần thông tin quan trọng ra khỏi lượng lớn thông tin được tạo dựa trên sở thích hoặc hành vi quan sát được của người dùng về một đối tượng item.

Hệ thống gợi ý có khả năng dự đoán liệu một người dùng cụ thể có thích một mặt hàng hay không dựa trên dữ liệu của người dùng trước

## Giới thiệu bài toán

---

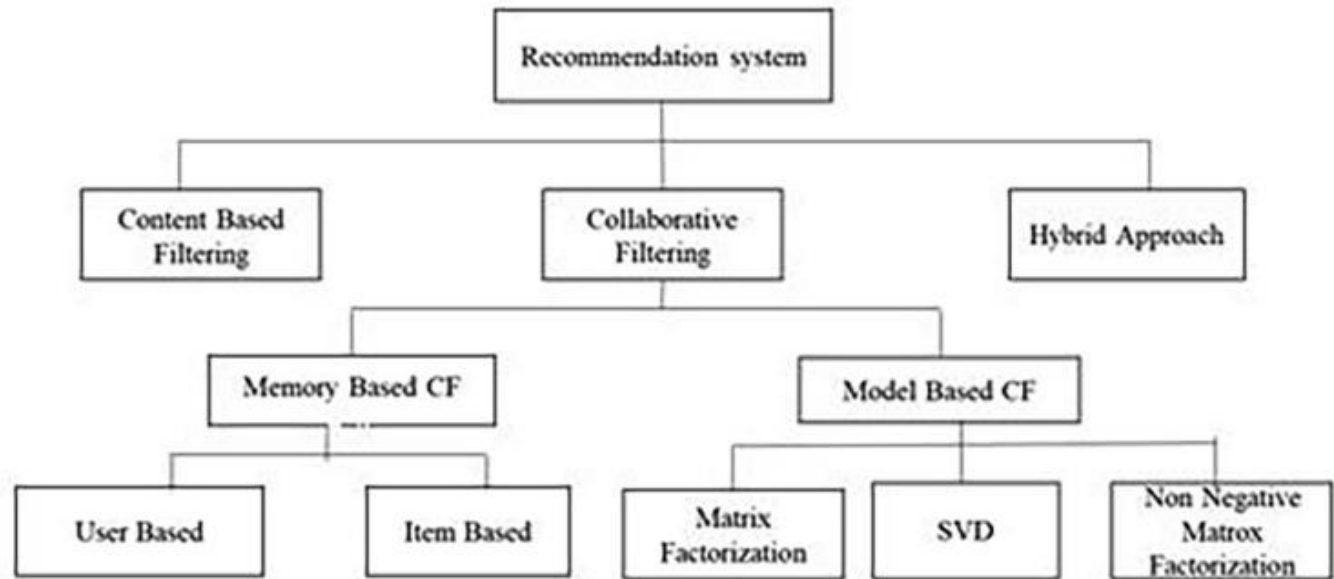
Hệ thống gợi ý được chia thành ba hướng tiếp cận:

1. Tiếp cận theo hướng lọc nội dung (Content-based filtering)
2. Tiếp cận theo hướng bộ lọc cộng tác (Collaborative filtering)
3. Phương pháp lai (Hybrid)

# Giới thiệu bài toán

I

Nội Dung  
Lý Thuyết



**Fig.1** Cây phân loại Recommendation system

Reference: N. Bhalse and R. J. M. T. P. Thakur, "Algorithm for movie recommendation system using collaborative filtering," 2021.

## Mục tiêu bài toán

---

Có hai thực thể chính quan trọng trong bài toán Recommendation System là người dùng (Customers) và sản phẩm (Items).

Items là sản phẩm có thể ví dụ như các bộ phim, bài hát, cuốn sách, video hoặc cũng có thể là customers khác trong bài toán gợi ý kết bạn.

Mục đích chính của các Recommender Systems là dự đoán mức độ quan tâm của một người tới một sản phẩm nào đó, quá đó có thể chọn chiến lược gợi ý phù hợp với cá nhân người dùng.



## Mục tiêu bài toán

---

Về cơ bản, mục đích của hệ thống đề xuất là tìm kiếm nội dung thú vị đối với một cá nhân. Hơn nữa, nó còn mở rộng một số yếu tố để tạo nên danh sách cá nhân hóa nội dung hữu ích và thú vị dành riêng cho từng cá nhân người dùng.

Một mục đích khác của hệ thống gợi ý là đạt được sự trải nghiệm của người dùng, giúp tối đa hóa thời gian của người dùng trên trang web hoặc kênh mà mình hướng đến, giúp tăng mức độ tương tác của khách hàng.



## II

### Phương Pháp Giải Quyết

# Phương Pháp Giải Quyết

---

Như đã phân loại hướng tiếp cận bài toán Recommendation System ở trên. Ta có thể chia phương pháp giải quyết bài toán Recommendation System thành ba hướng tiếp cận chính.

Tiếp cận theo hướng lọc nội dung (Content-based filtering), tiếp cận theo hướng bộ lọc cộng tác (Collaborative filtering) và dựa trên phương pháp lai (Hybrid)

## II

### Phương Pháp Giải Quyết

## Content-based filtering

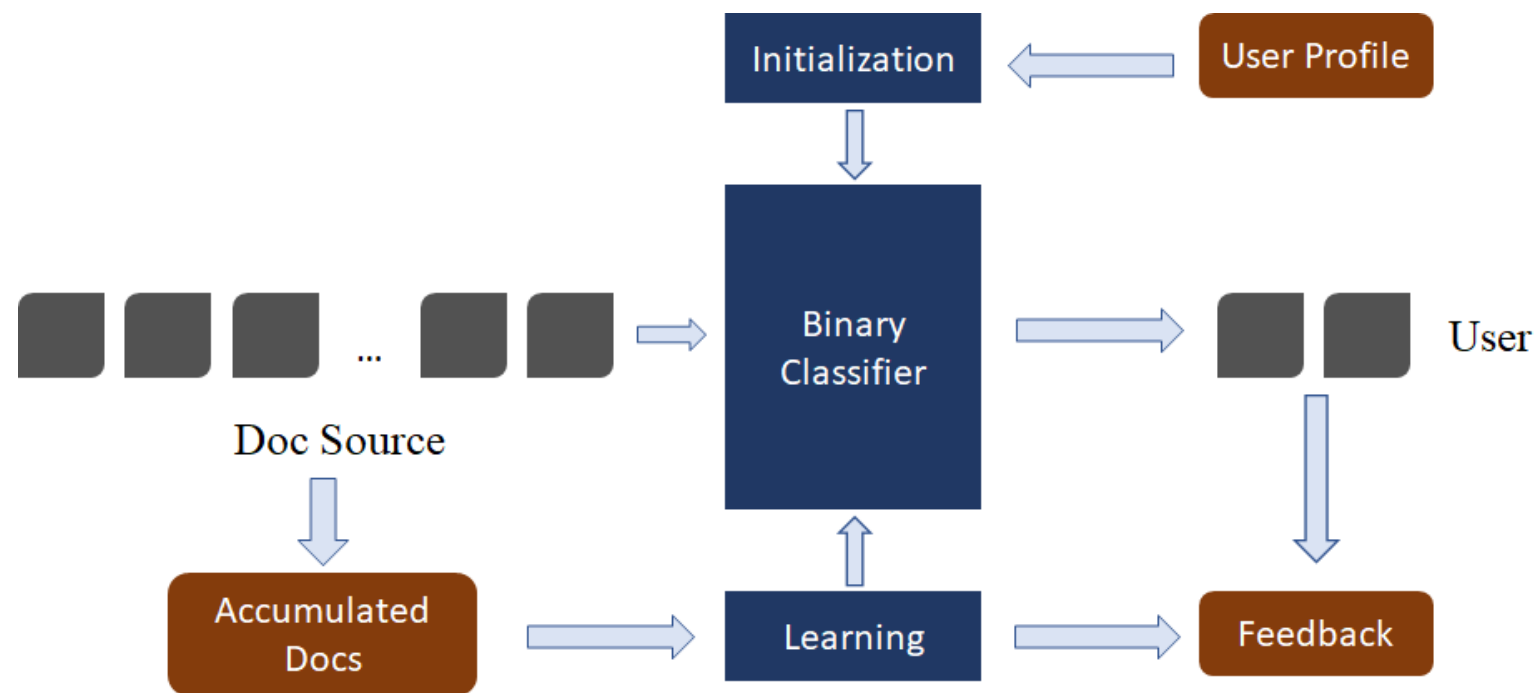
---

Phương pháp lọc dựa trên nội dung sử dụng các tính năng của mặt hàng để đề xuất các mặt hàng khác tương tự như những mặt hàng mà người dùng quan tâm, dựa trên các hành động trước đây của họ hoặc những phản hồi rõ ràng của người dùng như đánh giá 1-5 sao, tìm kiếm nhiều lần một nội dung hay dành nhiều thời gian cho một mặt hàng nào đó.

## II

## Phương Pháp Giải Quyết

# Content-based filtering



**Fig.2** Sơ đồ một hệ thống lọc dựa trên nội dung (Content-based filtering) điển hình

## II

### Phương Pháp Giải Quyết

# Content-based filtering

---

**Tạo bộ lọc ra quyết định Filtering Decision** (Binary Classifier): Bộ lọc đưa ra quyết định sẽ phân loại binary (yes/no) khi nhận đầu vào là dữ liệu text và mô tả hồ sơ của người dùng.

**Initialization Module:** Một bộ lọc khởi tạo ban đầu chỉ dựa trên profile text

**Learning:** Dựa trên những đánh giá đáng tin cậy (chỉ những .doc được đánh giá 'yes') và dựa trên các dữ liệu tích lũy được (Accumulated Docs) trong quá khứ

## II

### Phương Pháp Giải Quyết

# Content-based filtering

---

Cây quyết định (Decision trees), mạng Neural (Neural Network) và biểu diễn dựa theo Vector (Vector-based representations) đều có thể được sử dụng để học hồ sơ người dùng. Hồ sơ người dùng trong hệ thống lọc dựa trên nội dung là những dữ liệu lâu dài và được cập nhật theo thời gian

## II

### Phương Pháp Giải Quyết

# Ưu điểm Content-based filtering

---

- Không cần dữ liệu từ những người dùng khác để bắt đầu đưa ra các đề xuất.
- Các đề xuất dựa trên nội dung có thể được điều chỉnh phù hợp với sở thích của người dùng.
- Hệ thống lọc dựa trên nội dung thường dễ tạo hơn.

## II

### Phương Pháp Giải Quyết

# Thách thức Content-based filtering

---

- Hệ thống thiếu tính mới và đa dạng.
- Khả năng mở rộng mà một thách thức của hệ thống lọc dựa trên nội dung.
- Các thuộc tính có thể không chính xác hoặc không nhất quán.



## II

### Phương Pháp Giải Quyết

# Collaborative filtering

---

Phương pháp tiếp cận lọc theo nội dung chủ yếu phụ thuộc vào người dùng và thông tin của mặt hàng.

Hệ thống gợi ý dựa trên phương pháp lọc theo nội dung đề xuất một mặt hàng cho người dùng bằng cách phân tích hồ sơ (profile) của người dùng, thông tin của mặt hàng và phản hồi của người dùng (user feedback).

Collaborative filtering (CF) được chia thành hai hướng tiếp cận: Model-based CF và Memory-based CF

## II

### Phương Pháp Giải Quyết

## Memory-based CF

---

Memory-based CF sử dụng ma trận đánh giá để thực hiện dự đoán và gợi ý tư vấn.

Dựa trên tính toán sự tương đồng, Model-based CF có thể được phân loại thêm thành User-based CF (Lọc cộng tác dựa trên người dùng) và Item-based CF (Lọc cộng tác dựa trên sản phẩm). User-based CF sử dụng sự tương đồng giữa những người dùng trong khi đó Item-based CF xem xét hệ thống lọc cộng tác dựa trên sản phẩm

## II

### Phương Pháp Giải Quyết

# Memory-based CF

---

## Các bước thực hiện Memory-based CF:

1. Similarity Computation: Tính toán độ tương đồng giữa người dùng trong User-based CF và giữa các sản phẩm trong Item-based CF. Để tính độ tương quan, ta có thể dùng *Khoảng cách Consine*, *Khoảng cách Kullback-leiber*, *AHP*,...
2. Prediction: Ước lượng hay dự đoán giá trị đánh giá (rating) của người dùng đối với một sản phẩm nào đó dựa trên thông tin có được từ bước 1

## II

### Phương Pháp Giải Quyết

## Model-based CF

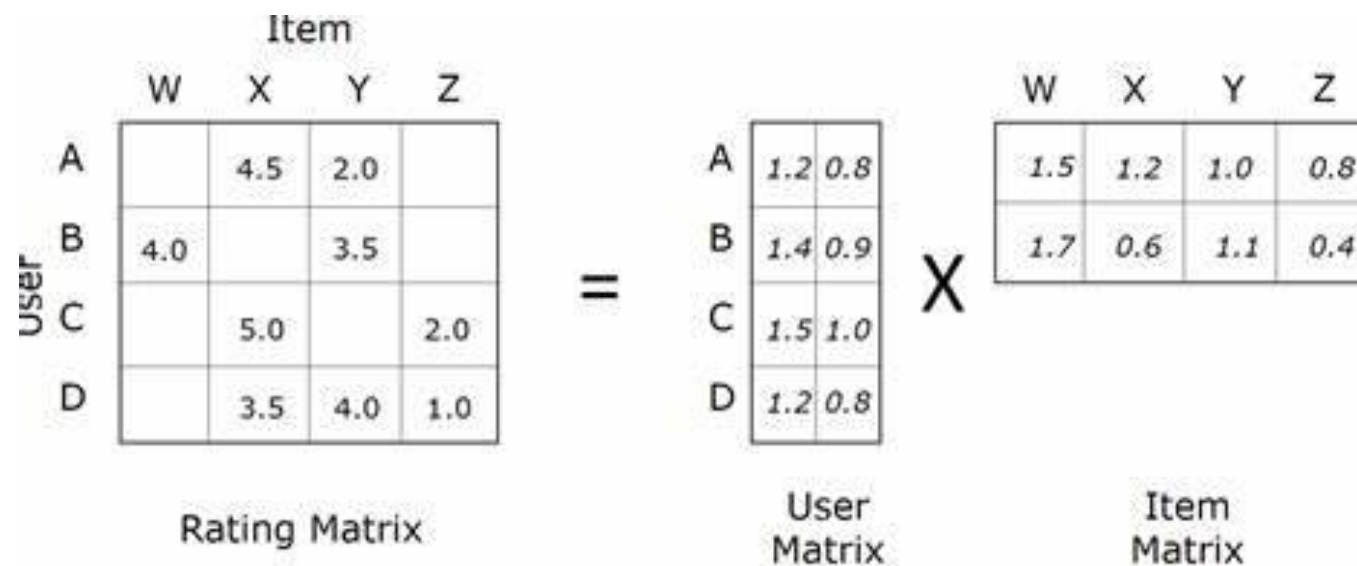
---

Model-based CF (Bộ lọc cộng tác dựa trên mô hình) yêu cầu các kỹ thuật học máy như phân loại (Classification), phân cụm (Clustering), hệ thống dựa trên quy tắc (Rule-based System),... để phát triển một mô hình được tối ưu hóa từ một bộ dữ liệu xếp hạng đã được thu thập từ người dùng

## II

# Phương Pháp Giải Quyết

## Model-based CF



**Fig.3** Sử dụng Matrix Factorization trong Recommender Systems

## II

### Phương Pháp Giải Quyết

# Ưu điểm Collaborative filtering

---

- Phương pháp gợi ý dựa trên bộ lọc cộng tác có nhiều lợi thế hơn so với lọc dựa trên nội dung.
- Hệ thống dựa trên collaborative filtering có khả năng nắm bắt sự thay đổi sở thích của người dùng theo thời gian, mô hình giúp người dùng khám phá những sở thích mới.
- Ở một mức độ nào đó, hệ thống chỉ cần ma trận phản hồi để đào tạo mô hình phân tích nhân tử ma trận.

## II

### Phương Pháp Giải Quyết

# Hybrid

---

Phương pháp lai (Hybrid) hay còn gọi là phương pháp kết hợp, Recommendation System sử dụng phương pháp lai sẽ kế thừa và kết hợp hai phương pháp giải quyết Content-based filtering và Collaborative filtering



## II

### Phương Pháp Giải Quyết

# Hybrid

---

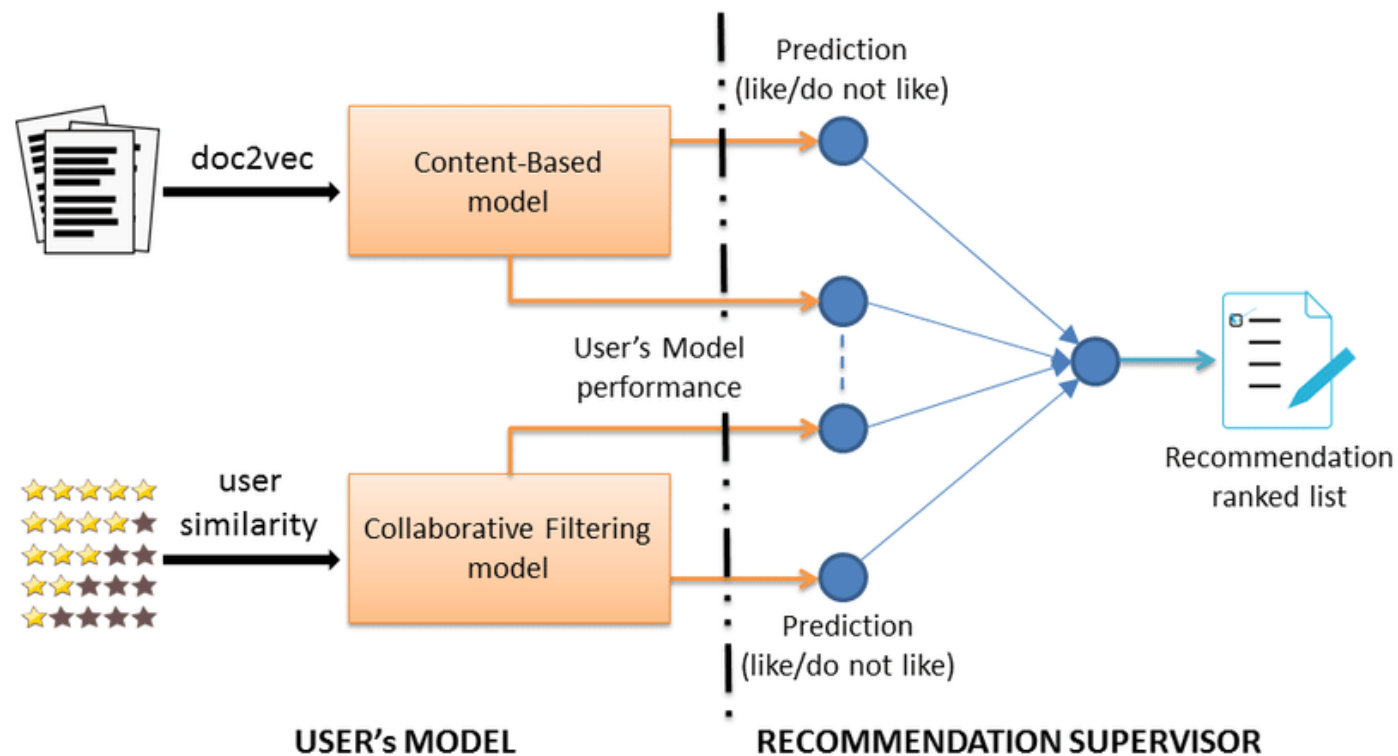
Sự kết hợp được thực hiện trong các phương pháp lai có thể được chia làm hai dạng:

1. Huấn luyện hai mô hình một cách độc lập (một mô hình lọc cộng tác và một mô hình dựa trên nội dung) và kết hợp các đề xuất của chúng
2. Trực tiếp xây dựng một mô hình để thống nhất cả hai cách tiếp cận bằng cách sử dụng làm thông tin trước khi nhập (về người dùng và/ hoặc vật phẩm) cũng như thông tin tương tác của cộng tác trực tuyến

## II

# Phương Pháp Giải Quyết

## Hybrid



**Fig.4** Mô hình Recommendation System dựa trên phương pháp lai.

## II

### Phương Pháp Giải Quyết

# Ưu điểm phương pháp Hybrid

---

Bởi vì kết hợp những ưu điểm và nhược điểm của bộ lọc nội dung và bộ lọc cộng tác, phương pháp này không chỉ tận dụng lợi thế của lọc nội dung mà còn có thể thực hiện lọc đối chiếu so sánh tương tự cho tất cả các sản phẩm, đặc biệt khi các sản phẩm không được đánh giá bởi bất kỳ người dùng nào và có thể được lọc ra, đề xuất cho người dùng, tránh vấn đề thiếu thông tin

## II

### Phương Pháp Giải Quyết

# Ưu điểm phương pháp Hybrid

---

Đồng thời phương pháp này cũng tận dụng được những ưu điểm của phương pháp lọc cộng tác. Khi số lượng người dùng và mức độ đánh giá tương đối lớn, ma trận dữ liệu xếp hạng người dùng của bộ lọc cộng tác sẽ trở nên dày đặc hơn, tăng số lượng chiều của ma trận.

Phương pháp này có thể làm giảm độ thừa thớt của ma trận và làm cho lọc cộng tác chính xác hơn.



CẢM ƠN THẦY CÔ VÀ CÁC BẠN  
ĐÃ LẮNG NGHE

---

