

Đồ án chuyên ngành

Khảo sát hệ thống khuyến nghị trên một số trang web xem phim

Nội dung chính

1

Giới thiệu về hệ thống khuyến nghị

- ❖ Giới thiệu, Lý thuyết bài toán khuyến nghị
- ❖ Các hướng tiếp cận

2

Một số phương pháp trong việc giải quyết bài toán khuyến nghị phim

- ❖ PMF (Probabilistic Matrix Factorization)
- ❖ BPMF (Bayesian Probabilistic Matrix Factorization)
- ❖ ALS (Alternating Least Squares)

3

Khảo sát hệ thống khuyến nghị trên một số trang web xem phim

- ❖ Netflix
- ❖ FPT Play và Zing TV

PHẦN 1

Giới thiệu về
hệ thống khuyến nghị

Giới thiệu về hệ thống khuyến nghị

Hệ thống khuyến nghị (Recommender System) hay còn được gọi là hệ thống tư vấn, là một hệ thống có nhiệm vụ chọn lọc thông tin nhằm dự đoán sở thích, mức độ phù hợp, mối quan tâm và nhu cầu của người dùng để đưa ra một hoặc nhiều mục, sản phẩm, dịch vụ mà người dùng sẽ quan tâm với xác suất là lớn nhất

Lý thuyết bài toán khuyến nghị

Lý thuyết bài toán khuyến nghị bao gồm các định nghĩa như:

- ❖ Không gian người dùng
- ❖ Không gian đối đối tượng khuyến nghị
- ❖ Hàm phù hợp

Lý thuyết bài toán khuyến nghị

Trong đó:

- ❖ **Không gian người dùng** là tập tất cả những người dùng mà hệ thống quan sát được.
- ❖ **Không gian đối đối tượng khuyến nghị** là tập tất cả những đối tượng sẽ được khuyến nghị cho người dùng. Ví dụ: Mặt hàng, phim ảnh, sách vở...
- ❖ **Hàm phù hợp** là một ma trận ánh xạ mức độ phù hợp của người dùng đối với đối tượng khuyến nghị

Các hướng tiếp cận của bài toán khuyến nghị

Các hướng tiếp cận bao gồm

- ❖ Phương pháp lọc dựa trên nội dung (Content-base Filtering)
- ❖ Phương pháp lọc cộng tác (Collaborative Filtering)
- ❖ Phương pháp tiếp cận lai (Hybird Filtering)

Các hướng tiếp cận của bài toán khuyến nghị

Phương pháp lọc dựa trên nội dung (Content-based Filtering) là phương pháp thực hiện dựa trên việc so sánh nội dung thông tin hay mô tả hàng hóa để tìm ra những sản phẩm tương tự với những gì mà người dùng đã từng quan tâm và khuyến nghị cho họ những sản phẩm này.

❖ Các cách tiếp cận lọc dựa trên nội dung:

- Tiếp cận dựa trên bộ nhớ
- Tiếp cận dựa trên mô hình

Các hướng tiếp cận của bài toán khuyến nghị

Phương pháp lọc cộng tác (Collaborative Filtering) là phương pháp khai thác những khía cạnh liên quan đến thói quen sử dụng sản phẩm của một nhóm người dùng có cùng sở thích trong quá khứ để đưa ra dự đoán các sản phẩm mới phù hợp với người dùng hiện tại.

❖ Các cách tiếp cận lọc cộng tác:

- Tiếp cận dựa trên bộ nhớ
- Tiếp cận dựa trên mô hình

Các hướng tiếp cận của bài toán khuyến nghị

Phương pháp tiếp cận lai (Hybird Filtering) là phương pháp kết hợp các kỹ thuật khuyến nghị khác nhau. Ví dụ như lọc cộng tác kết hợp với lọc dựa trên nội dung và ngược lại.

❖ Một số phương pháp tiếp cận lai:

Lai có trọng số

Lai chuyển đổi

Lai trộn

Lai meta

Lai kết hợp đặc trưng

Lai theo đợt

Lai tăng cường đặc trưng

PHẦN 2

Một số phương pháp
trong việc giải quyết
bài toán khuyến nghị phim

PMF (Probabilistic Matrix Factorization)

- ❖ Thuật toán PMF là một phương pháp trong hệ thống gợi ý dựa trên kỹ thuật phân rã ma trận
- ❖ Thuật toán PMF dựa trên hướng tiếp cận lọc cộng tác
- ❖ Thuật toán PMF được sử dụng rộng rãi trong các hệ thống gợi ý
- ❖ Hoạt động bằng cách tìm cách phân rã ma trận người dùng-sản phẩm thành hai ma trận nhỏ hơn, mỗi ma trận biểu diễn cho người dùng và sản phẩm
- ❖ PMF có hạn chế như việc không thể xử lý dữ liệu bị thiếu và không thể tự động điều chỉnh các tham số mô hình.

BPMF (Bayesian Probabilistic Matrix Factorization)

- ❖ Thuật toán BPMF là một phương pháp tiếp cận Bayesian đầy đủ của mô hình PMF.
- ❖ BPMF có thể được huấn luyện hiệu quả bằng cách sử dụng các phương pháp Markov chain Monte Carlo.
- ❖ Kết quả thu được từ các mô hình BPMF đạt độ chính xác dự đoán cao hơn đáng kể so với các mô hình PMF được huấn luyện bằng cách sử dụng ước lượng MAP
- ❖ Ưu điểm quan trọng của BPMF là tránh được việc điều chỉnh tham số và cung cấp phân phối dự đoán

ALS (Alternating Least Squares)

- ❖ Thuật toán ALS là một phương pháp được sử dụng trong hệ thống gợi ý dựa trên kỹ thuật phân rã ma trận
- ❖ Thuật toán ALS là một phương pháp phổ biến trong lọc cộng tác
- ❖ Thuật toán ALS được sử dụng để giải quyết vấn đề quá khớp (overfitting) trong dữ liệu thưa thớt và tăng độ chính xác của dự đoán trong hệ thống gợi ý
- ❖ Ưu điểm của thuật toán ALS là nó có thể xử lý dữ liệu bị thiếu và có thể mở rộng để xử lý dữ liệu lớn.

PHẦN 3

Khảo sát hệ thống
khuyến nghị trên một số
trang web xem phim

Netflix

Đánh giá hệ thống khuyến nghị của Netflix qua quá trình khảo sát:

- ❖ Do là một trang web xem phim lớn với độ bao phủ toàn cầu nên hệ thống khuyến nghị của trang web xem phim được xây dựng rất tốt.
- ❖ Có khả năng thu thập và xây dựng bộ hồ sơ người dùng thông qua quá trình người dùng tương tác trang web.
- ❖ Khả năng xây dựng hồ sơ người dùng ngay từ lúc tạo tài khoản người dùng.
- ❖ Có thể khuyến nghị các bộ phim mới, đang hot hoặc các bộ phim có độ tương đồng cao cho người dùng trong quá trình xem phim.
- ❖ Có khả năng mở rộng sở thích của người dùng bằng cách đưa ra các bộ phim mà có khả năng phù hợp với người dùng nhất.

FPT Play và Zing TV

Đánh giá hệ thống khuyến nghị của FPT Play và Zing TV qua quá trình khảo sát:

- ❖ Có khả năng hiển thị các bộ phim hot gần đây, nhằm hướng tới sở thích cá nhân của người dùng phù hợp.
- ❖ Hệ thống khuyến nghị được thiết kế đơn giản, tập trung vào hiệu suất. Do đó chất lượng của hệ thống khuyến nghị chỉ dừng ở mức cơ bản.
- ❖ Mặc dù hệ thống của FPT Play có lưu trữ tài khoản người dùng nhưng lại không có khả năng xây dựng bộ hồ sơ người dùng nhằm đưa ra khuyến nghị cá nhân. Còn của Zing TV thì không lưu trữ tài khoản người dùng.
- ❖ Khả năng khuyến nghị của trang web xem phim FPT Play và Zing TV chỉ dừng lại ở mức khuyến nghị dựa trên nội dung của các bộ phim. Ví dụ: khuyến nghị các bộ phim dựa theo *thể loại, nội dung, diễn viên....*

Nhận định chung sau quá trình khảo sát

Sau quá trình khảo sát, em đưa ra nhận định như rằng hầu hết các trang web xem phim hiện hay đều chia làm 2 dạng hệ thống thống khuyến nghị tùy thuộc vào dịch vụ mà trang web đó hướng tới:

- ❖ Dạng 1: Có xây dựng hồ sơ người dùng. Ví dụ: Netflix
- ❖ Dạng 2: Không xây dựng hồ sơ người dùng. Ví dụ: FPT Play và Zing TV

=> Đối với dạng 1 thường là những trang web xem phim lớn, được xây dựng với mục đích đáp ứng lượng người dùng lớn và mang lại trải nghiệm tốt cho người dùng. Do đó, việc hệ thống khuyến nghị có khả năng xây dựng bộ hồ sơ người dùng sẽ giúp cho trải nghiệm người dùng tốt hơn, các hệ thống nhờ đó mà thu thập dữ liệu để học tập và phát triển các mô hình dự đoán khuyến nghị tốt hơn nhằm mang lại chất lượng khuyến nghị tốt nhất.

=> Đối với dạng 2 thường là những trang web xem phim nhỏ, được xây dựng đơn giản nhằm tối ưu hóa chi phí phát triển cùng với đó mục đích thu lợi nhuận thông qua quảng cáo trên trang web khi người dùng xem phim trực tuyến trên trang web đó. Do đó, hệ thống khuyến nghị trên những trang web như thế thường dừng ở mức cơ bản.