

NGHIÊN CỨU VÀ ỨNG DỤNG PHƯƠNG PHÁP COLLABORATIVE FILTERING VÀ CONTENT-BASED FILTERING TRONG XÂY DỰNG MÔ HÌNH KHUYẾN NGHỊ PHIM



Nguyễn Ngọc Long - KHMT01 - K15

Mô tả vắn tắt

Đồ án này sử dụng hai phương pháp phổ biến trong lĩnh vực Data Mining để xây dựng mô hình gợi ý phim là Collaborative Filtering hay còn gọi là Lọc cộng tác và Content-Based Filtering hay còn gọi là Lọc theo nội dung.

THÔNG TIN BỘ DỮ LIỆU

Đồ án này sử dụng bộ dữ liệu MovieLens - 1M chứa thông tin của 1.000.029 đánh giá ẩn danh của khoảng 3900 phim tạo bởi 6.040 người dùng MovieLens vào năm 2000. Bao gồm 3 tệp dữ liệu: Movies.dat, Users.dat, Ratings.dat.

GIẢI THUYẾT

Vận dụng kiến thức của hai phương pháp được sử dụng phổ biến trong các bài toán hệ thống gợi ý là Collaborative Filtering kết hợp cùng giải thuật K-Nearest Neighbors và Content-Based Filtering để xây dựng mô hình gợi ý.

THÍ NGHIỆM

Đồ án này nhằm mục đích xây dựng một mô hình gợi ý phim và giúp cho các nhà quản lý của các website xem phim trực tuyến đưa ra gợi ý thông qua dữ liệu về người dùng hoặc tên phim, hoặc cũng có thể kết hợp cả hai để đưa ra những bộ phim có thể những người dùng này sẽ thích nhằm mục đích cải thiện sự nổi tiếng của các trang website của họ hoặc tăng doanh thu.

Movie Recommendations

Bộ dữ liệu được sử dụng trong bài báo cáo là bộ dữ liệu movie-lens 1M DataSet được tạo ra bởi GroupLens Research là một nhóm nghiên cứu thuộc Đại học Minnesota và được phát hành vào tháng 2 năm 2003.

Bộ dữ liệu này chứa thông tin của 1.000.209 đánh giá ẩn danh của khoảng 3.900 phim được tạo bởi 6.040 người dùng MovieLens vào năm 2000. Bao gồm 3 tệp dữ liệu: movies.dat chứa thông tin của các bộ phim bao gồm: MovieId, Title(tên phim) và Genres(thể loại phim), users.dat chứa thông tin của UserId, Giới tính, Tuổi, Nghề nghiệp, mã zip, Ratings.dat chứa thông tin của UserId, MovieId, Rating(số điểm đánh giá) và Timestamp(Thời gian đánh giá).

Select User ID for Collaborative Filtering

1

Gợi ý Collaborative

Select Movie for Content-Based Filtering

Toy Story (1995)

Gợi ý Content-Based

Select User ID for Hybrid Filtering

1

Select Movie for Hybrid Filtering

Toy Story (1995)

Gợi ý Hybrid

KẾT QUẢ VÀ KẾT LUẬN

Mô hình gợi ý phim theo sở thích được ứng dụng rộng rãi trong các trang web phim ảnh nổi tiếng như Netflix.com, Amazon Prime Video, Hulu, Disney+, IMDb TV... và nhiều dịch vụ khác. Các trang web này sử dụng các thuật toán và phương pháp khác nhau để gợi ý phim dựa trên sở thích và hành vi xem phim của người dùng.