Họ và tên: Ngô Minh Toàn

MSSV: 51801031

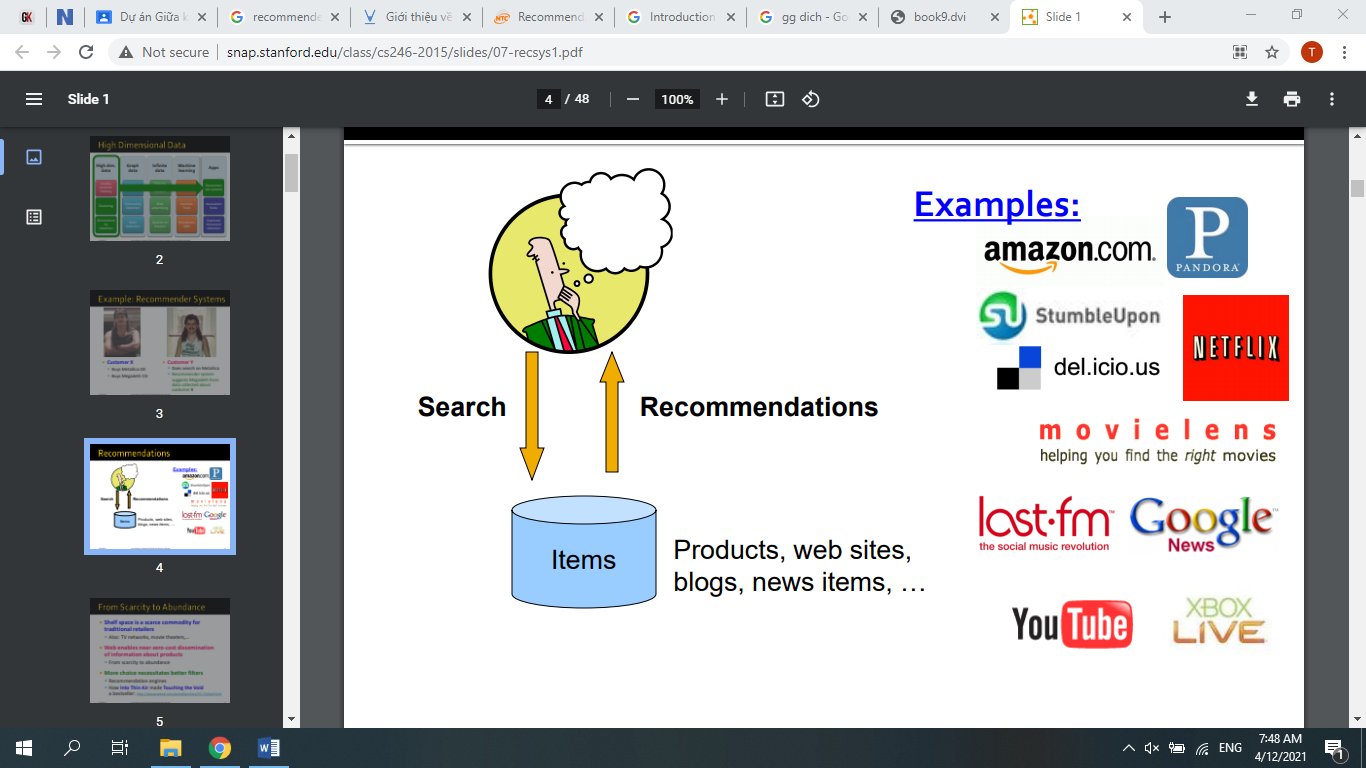
Lớp: 18050402

Tìm hiểu về bài toán Recommender Systems

1. Mục tiêu của recommender systems

Hệ thống gợi ý (Recommender systems hoặc Recommendation systems) là các thuật toán nhằm mục đích đề xuất các mục có liên quan cho người dùng, hỗ trợ ra quyết định, cung cấp giải pháp mà không phải trải qua quá trình tìm kiếm phức tạp.

Với sự nổi lên của Youtube, Amazon, Netflix và nhiều dịch vụ web khác, các hệ thống giới thiệu ngày càng có nhiều vị trí trong cuộc sống của chúng ta. Từ thương mại điện tử (gợi ý cho người mua những bài báo mà họ có thể quan tâm dựa trên dự đoán sở thích của họ) đến quảng cáo trực tuyến (gợi ý cho người dùng những nội dung phù hợp dựa trên lịch sử mua hàng, hoặc sở thích của họ), hệ thống giới thiệu ngày nay không thể tránh khỏi trong việc trực tuyến hàng ngày của chúng ta.



2. Trình bày bài toán, mô hình giải quyết bài toán

Mục đích của hệ thống giới thiệu là đề xuất các mặt hàng có liên quan cho người dùng. Để đạt được nhiệm vụ này, tồn tại hai loại phương pháp chính: phương pháp lọc cộng tác (collaborative filtering) và phương pháp dựa trên nội dung (content based). Hãy thảo luận ngắn gọn về hai mô hình chính này.

2.1 Collaborative filtering

Phương pháp lọc cộng tác (collaboractive filtering) là phương pháp chỉ dựa trên các tương tác trong quá khứ được ghi lại giữa người dùng và các mặt hàng để đưa ra các đề xuất mới, đây là phương pháp gợi ý được triển khai rộng rãi nhất và thành công nhất trong thực tế. Những tương tác này được lưu trữ trong "ma trận user-item interactions". Sau đó sử dụng các tương tác giữa người dùng với các mục item này để phát hiện những người dùng tương tự và các item tương tự và đưa ra dự đoán dựa trên những khoảng cách ước tính này. Cho I là tập các đối tượng (Item) có thể được gợi ý, X là tập người dùng, u là một người dùng cụ thể trong tập X, và i là một đối tượng cụ thể trong I.

* Input: Các điểm số đánh giá của những người sử dụng trong X đối với các đối tượng trong I.
* Output: Các điểm số đánh giá của u cho các đối tượng trong I.
* Tiến trình xử lý: Nhận ra người sử dụng trong X tương tự với u ( về sở thích) và sau đó ngoại suy điểm số đánh giá của u cho i.

Thế mạnh lớn nhất của kỹ thuật collaboractive filtering là chúng hoàn toàn độc lập với sự biểu diễn của các đối tượng đang được gợi ý, có thể làm việc tốt với các đối tượng phức tạp như âm thanh và phim. Chất lượng của kỹ thuật này phụ thuộc vào độ lớn dữ liệu lịch sử tương tác, thao tác của người dùng.

2.2 Content based

Hệ thống gợi ý dựa theo nội dung (content based) là sự kế thừa và mở rộng của lĩnh vực nghiên cứu lọc thông tin. Không giống như các phương pháp cộng tác chỉ dựa vào tương tác giữa người dùng với item, các phương pháp dựa trên nội dung sử dụng thông tin bổ sung về người dùng hoặc mục. Trong hệ thống thì các đối tượng được biểu diễn bởi các đặc điểm liên quan tới chúng.

Ý tưởng của các phương pháp dựa trên nội dung là cố gắng xây dựng một mô hình, dựa trên các “tính năng” có sẵn, giải thích các tương tác giữa người dùng và mục được quan sát. Ví dụ để tạo thành một model xem xét người dùng và phim, ta thử mô hình hóa thực tế là phụ nữ trẻ có xu hướng đánh giá tốt hơn một số phim, nam giới trẻ có xu hướng đánh giá tốt hơn một số phim khác, v.v. Nếu ta có được mô hình như vậy, thì việc đưa ra các dự đoán mới cho người dùng khá dễ dàng bằng cách chỉ cần xem hồ sơ (tuổi, giới tính,…) của người dùng này và dựa trên thông tin này để xác định các bộ phim có liên quan để đề xuất.

Cho I là tập các đối tượng (Item) có thể được gợi ý, X là tập người dùng, u là một người dùng cụ thể trong tập X, và i là một đối tượng cụ thể trong I.

* Input: Các đặc điểm của các đối tượng trong I.
* Output: Các điểm số đánh giá của u cho các đối tượng trong I.
* Tiến trình xử lý: Tạo ra một mô hình mô tả sở thích của người sử dụng u, sau đó sử dụng để đánh giá mức độ ưa thích của u với i.

Cũng giống như phương pháp lọc cộng tác, hồ sơ người dùng trong gợi ý dựa theo nội dung là những dữ liệu lâu dài và được cập nhật theo thời gian. Chất lượng phụ thuộc vào độ lớn dữ liệu. Các phương pháp dựa trên nội dung ít gặp phải vấn đề khởi đầu hơn so với các phương pháp cộng tác: người dùng hoặc mục mới có thể được mô tả theo đặc điểm của họ (nội dung) và do đó, các đề xuất phù hợp có thể được thực hiện.

TÀI LIỆU KHAM THẢO

1. <http://snap.stanford.edu/class/cs246-2015/slides/07-recsys1.pdf>
2. <http://infolab.stanford.edu/~ullman/mmds/ch9.pdf>
3. <https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada>
4. <https://viblo.asia/p/gioi-thieu-ve-he-thong-goi-y-recommender-systems-hoac-recommendation-systems-maGK78yOZj2>