dòng ngang

**Vuong Tuan Khanh**

Recommender System

## Mô hình của Amazon

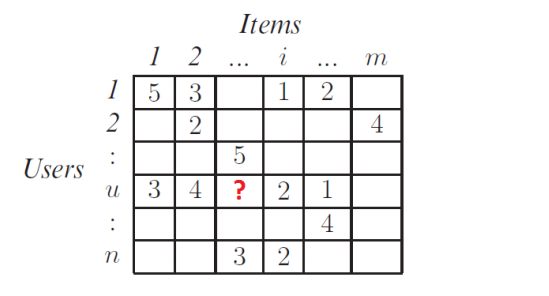
* Quan tâm đến việc khách hàng yêu thích những sản phẩm nào dựa vào dữ liệu trên quá khứ của họ như điểm đánh giá trên từng sản phẩm, thời gian duyệt trên từng sản phẩm, số lần click vào sản phẩm…
* Từ đó có thể dự đoán được người dùng có thể sẽ thích những sản phẩm nào khác và đưa ra gợi ý phù hợp cho họ

## Các thành phần cơ bản

* Cần phải quan tâm đến 3 đặc điểm cơ bản:
* User
* Item
* Feedback của mỗi User trên một Item
* Điểm đánh giá
* Chỉ số quan tâm

## Biểu diễn thông tin

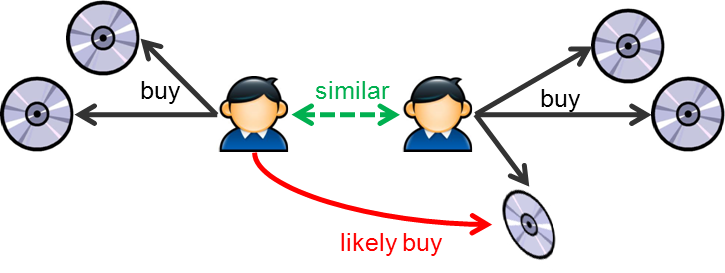
* Thông tin phải được biểu diễn dưới dạng tính toán được
* Ý tưởng:
* Dùng ma trận
* Ý nghĩa:
* Một ma trận được tạo ra thể hiện độ thích của từng user lên các item tương ứng



* + Những ô có trọng số sẽ thể hiện được mức độ yêu thích của mỗi user lên các item
  + Những ô còn trống thể hiện user chưa từng tiếp cận được với item
* Mục tiêu:
  + Dự đoán các giá trị tại các ô còn trống trong ma trận trên rồi sắp xếp theo thứ tự độ thích giảm dần để gợi ý cho người dùng

## Phân loại

* **Hệ thống gợi ý dựa trên nội dung – Content based recommender systems:**
* Hệ thống sẽ quan tâm đến nội dung, đặc điểm của mục tin hiện tại và sau đó gợi ý cho người dùng các mục tin tương tự.
* **Hệ thống gợi ý dựa trên các user – lọc cộng tác – Collaborative filtering recommender systems:**
* Hệ thống sẽ phân tích các user có cùng đánh giá, cùng mua mục tin hiện tại. Sau đó tìm ra danh sách các mục tin khác cũng được đánh giá bởi các user này, xếp hạng và gợi ý cho người dùng. Tư tưởng của phương pháp này chính là dựa trên sự tương đồng về sở thích giữa các người dùng để đưa ra các gợi ý.



* **Phụ trợ**:
* **Demographic**: Hệ thống này sẽ gợi ý item dựa nhân khẩu học của user. Giả thuyết cho rằng những thị trường khác nhau nên được gợi ý những item khác nhau. Chẳng hạn, user sẽ được điều hướng tới những website khác nhau dựa trên ngôn ngữ và địa lý. Hoặc là việc gợi ý có thể thay đổi dựa trên tuổi của user. Mặc dù cách tiếp cận này khá phổ biến trong mảng tiếp thị truyền thống, nhưng có rất ít nghiên cứu khoa học về kỹ thuật này.
* **Knowledge-based**: hệ thống sẽ gợi ý item dựa trên miền kiến thức cụ thể rằng những đặc điểm nào của item đáp ứng được nhu cầu và thị hiếu của user, cuối cùng là item đó có hữu ích cho user hay không.
* **Community-based**: gợi ý item dựa trên sở thích của bạn bè của user. Kết quả gợi ý sẽ dựa trên rating thu thập được từ bạn của user.
* **Hybrid recommender system**: Kết hợp của những kỹ thuật đã đề cập ở trên. Lấy ví dụ, phương pháp CF sẽ gặp vấn đề với các item mới, nghĩa là nó không thể sinh ra gợi ý cho các item mà chưa có rating. Tuy nhiên điều này lại không ảnh hưởng đến content-based vì phương pháp này dự đoán cho các item mới dựa trên mô tả của chúng (các tính năng, đặc điểm). Với hai kỹ thuật RSs cơ bản (hoặc nhiều hơn) kết hợp lại tạo ra một hệ thống hybrid mới nhằm tận dụng lợi thế của kỹ thuật này để bổ sung cho điểm yếu của kỹ thuật kia.

## Đánh giá

* **Phương pháp gợi ý dựa trên nội dung** đòi hỏi phải thu thập rất nhiều thông tin về các mục tin tương tự, đòi hỏi phải thu thập và phân tích, xử lý toàn bộ các mục tin trong cơ sở dữ liệu.
* **Phương pháp lọc cộng tác**: Không cần quá nhiều thông tin. Đơn giản chỉ là item\_id của item hiện tại, các user\_id và các feedback trên item đó mà thôi nên thực tế thì phương pháp lọc cộng tác được sử dụng phổ biến hơn để xây dựng các hệ thống gợi ý

## Quy trình

* **Thu thập Data**
* Rating mỗi Item
* Số lần click chuột vào item
* Thời gian trung bình thao tác với item
* etc
* **Chuẩn hóa**
* **Chạy mô hình, lọc ra top N item phù hợp**
* **Đánh giá mô hình**

## Lưu ý

* Cần bổ sung thêm một số luật để chọn ra được top N item phù hợp hơn với từng user.
* Không cho hiển thị những sản phẩm đã được mua trong danh sách gợi ý.
* Cần một kế hoạch cập nhật lại mô hình thông qua tập dữ liệu mới sinh ra hàng ngày.

## Những yêu cầu cơ bản

* Các yêu cầu của bài toán đặc thù, các chỉ tiêu, KPI, metrics đánh giá hiệu quả, các tính năng, thiết kế, hệ thống, module, công nghệ sử dụng
* Kết nối các thành phần, phân hệ, phòng ban, ngành hàng, phát triển sản phẩm, nhân sự
* Các thông tin cần thu thập để phân tích lựa chọn của user
* Các mô hình phổ biến và mô hình nâng cao hơn, cấu trúc, cách hoạt động, UX…
* Tóm gọn:
* Tăng năng suất
* Nắm và phán đoán được nhu cầu khách hàng