

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

KHAI PHÁ WEB ĐỀ TÀI: XÂY DỰNG HỆ GỢI Ý ÂM NHẠC

Giảng viên hướng dẫn: TS.Nguyễn Kiêm Hiếu

Sinh viên: Nguyễn Thị Dung 20140698

Lại Trung Kiên 20142398

Võ Văn Tài 20143927

Bùi Tiến Thành 20144052

Giới thiệu

Đúng như tên gọi của hệ thống, hệ gợi ý đưa ra những đề xuất những sản phẩm cho người dùng mà được hệ thống đánh giá là phù hợp hợp nhất.

Có thể bạn không để ý nhưng những trang web lớn trên thế giới đa số đều tích hợp một hệ gợi ý thông minh vào trong hệ thống của họ.

- Youtube đưa ra danh sách những video theo chủ đề ở phần trang chủ, hay tự chuyển khi kết thúc một video.
- Facebook đưa ra gợi ý những bạn bè mà có thể bạn biết.
- Amazon gợi ý những sản phẩm bạn có thể muốn mua cùng với sản phẩm bạn vừa mới đặt hàng...

Đó đều là những ví dụ điển hình mà một hệ gợi ý thông minh có thể mang lại.

Giới thiệu

Vậy thuật toán đằng sau những hệ thống gợi ý này là gì? Và mục đích chính của hệ gợi ý này là ntn?

- Là những giải thuật Machine Learning có tên gọi chung là Recommendation Systems.
- Trong Recommendation Systems có hai thực thế chính đó là user và item
- Mục đích chính là dự đoán mức độ quan tâm của người dùng tới một số item nào đó, qua đó đưa ra chiến lược gợi ý phù hợp

Một số phương pháp trong hệ gợi ý

- Hệ gợi ý dựa theo nội dung (Content-based systems)
- Hệ gợi ý dựa theo lọc cộng tác (Collaborative filtering)
- Hệ gợi ý dựa theo cơ sở tri thức (Knowledge-based systems)

- Hệ gợi ý dựa theo nội dung (Content-basead systems)
 - Đưa ra gợi ý cho người dùng dựa trên mô tả với những đặc điểm của item và hồ sơ về sở thích của người dùng trước đó
- Hệ gợi ý dựa theo cơ sở tri thức (Knowledge-based systems)
 - Là hệ thống gợi ý các đối tượng dựa trên các suy luận về nhu cầu và sở thích của người dùng. Theo một nghĩa nào đó, tất cả các kỹ thuật gợi ý có thể mô tả như là làm một số suy luận.

- Hệ gợi ý dựa theo lọc cộng tác (Collaborative filtering)
 - Là phương pháp gợi ý được triển khai rộng rãi nhất và thành công nhất trong thực tế.
 - Hệ thống theo lọc công tác phân tích và tổng hợp các điểm số đánh giá của các đối tượng, nhận ra sự tương đồng giữa những người sử dụng trên cơ sở các điểm số đánh giá của họ và tạo ra các gợi ý dựa trên sự so sánh này.

Hê gơi ý âm nhạc sử dụng phương pháp Collaborative filtering

Giới thiệu bài toán

- Mục đích: Xây dựng một hệ thống gợi ý bài hát cho người dùng dựa trên lịch sử nghe trước đó.
- Phương pháp: Sử dụng phương pháp Collaborative Filtering(CF) có tên là Neighborhood-based Collaborative Filtering(NBCF)
- Vấn đề quan trọng nhất của phương pháp này là trả lời hai câu hỏi
 - Làm thế nào để xác định sự giống nhau giữa hai users?
 - Khi đã xác định được các users gần giống nhau rồi, làm thế nào dự đoán được mức độ quan tâm của user đó lên item?

Hê gơi ý âm nhạc sử dung phương pháp Collaborative filtering

User-user collaborative filtering

	u_0	u ₁	u ₂	u_3	u_4	u ₅
i_0	6	6	2	3		
i_1	3	3			1	
i ₂			2			2
i ₃	4			5		4
i ₄	3	2	4			5

Utility matrix Y

Hê gơi ý âm nhạc sử dụng phương pháp Collaborative filtering

- User-user collaborative filtering
 - Hàm xác định sự giống nhau giữa hai user (Similarity functions)

$$cosine_similarity(u_1, u_2) = cos(u_1, u_2) = \frac{u_1^T u_2}{||u_1||_2 . \, ||u_2||_2} \quad (1)$$

Dự đoán rating

$$\hat{y}_{i,u} = \frac{\sum_{u_j \in \mathcal{N}(u,i)} \bar{y}_{i,u_j} \text{sim}(u,u_j)}{\sum_{u_j \in \mathcal{N}(u,i)} |\text{sim}(u,u_j)|} \quad (2)$$

Hê gơi ý âm nhạc sử dung phương pháp Collaborative filtering

User-user collaborative filtering

- Một số hạn chế
 - Số lượng users luôn lớn hơn số lượng items rất nhiều. Kéo theo đó là Similarity matrixlà rất lớn với số phần tử phải lưu giữ là hơn 1 nửa của bình phương số lượng users (chú ý rằng ma trận này là đối xứng). Việc này, nhưa đã đề cập ở trên, khiến cho việc lưu trữ ma trận này trong nhiều trường hợp là không khả thi.
 - Ma trận Utility Y thường là rất thưa. Với số lượng users rất lớn so với số lượng items, rất nhiều cột của ma trận này sẽ rất thưa, tức chỉ có một vài phần tử khác 0. Cũng chính vì việc này, một khi user đó thay đổi rating hoặc rate thêm items, trung bình cộng các ratings cũng như vector chuẩn hoá tương ứng với user này thay đổi nhiều. Kéo theo đó, việc tính toán ma trận Similarity, vốn tốn nhiều bộ nhớ và thời gian, cũng cần được thực hiện lại.

Hê gơi ý âm nhạc sử dung phương pháp Collaborative filtering

- Item-item collaborative filtering
 - Một số ưu điểm
 - Vì số lượng items thường nhỏ hơn số lượng users, Similarity matrix trong trường hợp này cũng nhỏ hơn nhiều, thuận lợi cho việc lưu trữ và tính toán ở các bước sau.
 - Vì số lượng phần tử đã biết trong Utility matrix là như nhau nhưng số hàng (items) ít hơn số cột (users), nên trung bình, mỗi hàng của ma trận này sẽ có nhiều phần tử đã biết hơn số phần tử đã biết trong mỗi cột. Kéo theo đó, giá trị trung bình của mỗi hàng ít bị thay đổi hơn khi có thêm một vài ratings. Như vậy, việc cập nhật ma trận Similarity Matrix có thể được thực hiện ít thường xuyên hơn.

Demo chương trình