

1 Giới thiệu

Cùng với sự phát triển của thương mại điện tử cũng như tương tác trực tuyến của con người trong giai đoạn hiện nay, hệ thống gợi ý đã được nghiên cứu và ứng dụng một cách mạnh mẽ và mang lại lợi ích cho cả người cung cấp dịch vụ và người sử dụng dịch vụ. Ý tưởng đằng sau một hệ thống gợi ý là sử dụng các dữ liệu thu thập được từ người dùng và dự đoán, gợi ý cho người dùng những sản phẩm, tính năng, dịch vụ mà người dùng có thể thích, từ đó nâng cao được chất lượng dịch vụ và thu lại lợi nhuận. Đối tượng mà hệ thống gợi ý hướng đến được gọi là **user**, còn sản phẩm mà hệ thống đưa ra gợi ý gọi là **item**. Do đó, hệ thống gợi ý thường dựa trên tương tác trong quá khứ giữa *user* và *item* để tạo nên dự đoán trong tương lai. (Có một ngoại lệ là trường hợp của *knowledge-based recommender systems*, hệ thống gợi ý sẽ dựa vào những yêu cầu cụ thể của user thay vì dựa vào dữ liệu trong quá khứ của user để đưa ra dự đoán. Chúng ta sẽ nói đến *knowledge-based recommender systems* trong những bài sau)

2 Mục tiêu của hệ thống gợi ý

Trước khi bàn luận về mục tiêu của hệ thống gợi ý, chúng ta sẽ nói về một số mô hình khác nhau của hệ thống gợi ý có thể được xây dựng. Có hai mô hình chính:

- Prediction version (phiên bản dự đoán): Cách tiếp cận đầu tiên dùng để dự đoán giá trị rating cho một cặp user-item. Giả sử chúng ta có một tập dữ liệu về mức độ quan tâm của m người dùng (user) lên n sản phẩm (item), tương ứng với một ma trận $m \times n$, trong đó phần tử ở hàng i , cột j thể hiện giá trị rating của user thứ i lên item thứ j . Việc cần làm của chúng ta là lấp đầy các giá trị còn trống trong ma trận, hay nói cách khác, chúng ta sẽ dự đoán giá trị rating của user lên items mà user đó chưa rate. Bài toán này còn được gọi là bài toán hoàn thành ma trận, bởi vì ma trận ban đầu mà chúng ta có được là ma trận chưa đầy đủ giá trị, những giá trị còn lại sẽ được dự đoán bằng cách sử dụng các thuật toán.
- Ranking version (phiên bản xếp hạng): Trong thực tế, thực sự chúng ta không cần thiết phải dự đoán tất cả các giá trị ratings của user lên items để đưa ra gợi ý cho user. Thay vào đó, chúng ta chỉ cần gợi ý top-k items phù hợp nhất cho user, hoặc xác định top-k users phù hợp nhất với item. Thông thường, việc gợi ý k items cho user là phổ biến hơn, mặc dù phương pháp để giải quyết hai vấn đề này là hoàn toàn tương tự nhau.

Trong mô hình thứ hai, giá trị ratings dự đoán được là không quan trọng. Mô hình thứ nhất phổ biến hơn trong thực tế, bởi vì mô hình thứ hai có thể được giải quyết thông qua việc xử lý mô hình thứ nhất. Tuy nhiên, trong nhiều trường hợp, sẽ là dễ dàng hơn và tự nhiên hơn nếu thiết kế và xây dựng mô hình thứ hai một cách trực tiếp.

Trong kinh doanh, việc tăng doanh số bán hàng hay tăng lợi nhuận là nhiệm vụ quan trọng nhất của hệ thống gợi ý. Bằng cách sử dụng các thuật toán, hệ thống gợi ý mang những items thích hợp đến sự chú ý của users, điều này làm tăng doanh thu và lợi nhuận cho người kinh doanh. Để thực hiện được nhiệm vụ này, hệ thống gợi ý phải có một số mục tiêu hoạt động và tiêu chí kỹ thuật như sau:

- Relevance (Tính phù hợp): Một mục tiêu hoạt động hiển nhiên nhất của hệ thống gợi ý là gợi ý items thích hợp với users, do users thường sử dụng những items mà họ cảm thấy thú vị và phù hợp. Mặc dù sự thích hợp là mục tiêu chính của hệ thống gợi ý, nhưng nếu chỉ có mỗi mục tiêu này thì vẫn là chưa đủ, vì vậy chúng ta sẽ nói đến một số mục tiêu "hạng 2" sau đây, mặc dù nó không quan trọng bằng Tính phù hợp, nhưng nó vẫn mang một ý nghĩa quan trọng và ảnh hưởng khá nhiều đến một hệ thống gợi ý được xem là tốt.
- Novelty (Tính mới lạ): Hệ thống gợi ý thực sự có hiệu quả khi nó gợi ý những items mà users chưa từng thấy trong quá khứ. Ví dụ, một bộ phim phổ biến với một thể loại quen thuộc sẽ hiếm khi gây ấn tượng cho user. Việc lặp đi lặp lại gợi ý những items phổ biến cũng có thể dẫn tới sự giảm sút trong sự đa dạng trong việc "bán" items.
- Serendipity: Trong mục tiêu này, items được gợi ý là một items không ngờ tới đối với user, và vì vậy, đây có thể được coi là cách gợi ý mang yếu tố may mắn, trái ngược với các mục tiêu gợi ý trước đó. *Serendipity* khác với *Novelty* ở chỗ những gợi ý này thực sự gây ngạc nhiên với user, thay vì gợi ý những items mà user chưa thấy trước đó. Mục tiêu này có thể được sử dụng trong trường hợp mà user đó chỉ quen

thuộc với một loại items cụ thể, mà không biết đến những loại items khác mà có thể gây ngạc nhiên và thích thú đối với những loại items mới này. Không giống với *Novelty*, *Serendipity* tìm cách mở rộng sự gợi ý đối với user. Đây có thể coi là một chiến lược mang tính lâu dài, nhằm mang đến những sở thích mới, những mối quan tâm mới cho users, từ đó có thể nâng cao được doanh số và lợi nhuận.

- Mở rộng sự đa dạng trong việc gợi ý: Hệ thống gợi ý thường gợi ý một danh sách top-k items. Nếu chúng quá giống nhau, có thể xảy ra trường hợp mà user có thể sẽ không thích items nào cả, và việc đó hoàn toàn không có lợi. Mặt khác, khi hệ thống gợi ý cố gắng đưa ra gợi ý các items ở nhiều loại khác nhau, khả năng mà user thích ít nhất một item sẽ tăng lên đáng kể, giảm được rủi ro trong việc gây nhàm chán cho user với việc gợi ý lặp đi lặp lại những sản phẩm quá giống nhau.

Bên cạnh những mục tiêu đã đề cập ở phía trên, còn có những mục tiêu bổ sung khác cũng được sử dụng trong hệ thống gợi ý từ cả hai phía người sử dụng và người cung cấp dịch vụ. Về phía của người sử dụng (users), việc gợi ý những sản phẩm phù hợp với cá nhân có thể giúp cho users thoả mãn và cảm thấy việc gợi ý của bên cung cấp là thực sự có ích với mình, tạo nên sự trung thành của users và tăng doanh số bán hàng. Cùng với đó, về phía người cung cấp dịch vụ, hệ thống gợi ý cũng cho phép phân tích được nhu cầu của users và từ đó giúp nâng cao được trải nghiệm của người dùng.

3 Một số ví dụ về hệ thống gợi ý

3.1 GroupLens Recommender System

GroupLens được coi là hệ thống gợi ý tiên phong, nó được xây dựng để gợi ý cho Usenet news. Hệ thống sẽ thu thập những ratings của người đọc Usenet và sử dụng chúng để dự đoán liệu một người dùng có thích bài báo hay không trước khi họ đọc chúng. Trong hệ thống này, một số thuật toán "collaborative filtering" đã được sử dụng. Một số ý tưởng sau đó cũng được phát triển dựa theo hệ thống này, những ý tưởng đó áp dụng cho sách và phim, với bộ dữ liệu khá quen thuộc cho những người mới bắt đầu là BookLens và MovieLens.

3.2 Amazon Recommender System

Amazon cũng được coi là một trong những hệ thống gợi ý tiên phong, đặc biệt là trong thương mại và hệ thống gợi ý đã được sử dụng rất thành công ở Amazon. Hệ thống này sử dụng đánh giá của người dùng, hành vi của người dùng và lịch sử hoạt động của người dùng để làm dữ liệu dự đoán. Ratings ở Amazon có thang điểm từ 1 đến 5, còn lịch sử mua hàng và lịch sử hoạt động có thể dễ dàng được thu thập bằng khi mà user thực hiện đăng nhập trên trang web của Amazon. Những gợi ý cho người dùng được hiển thị trên trang chủ của Amazon, ngay khi người dùng truy cập vào trang web. Trong nhiều trường hợp, những items được gợi ý cũng được giải thích, ví dụ như trong mối quan hệ của sản phẩm được gợi ý với những sản phẩm mà user đã mua trước đó.

3.3 Facebook Friend Recommendations

Facebook thường gợi ý cho chúng ta kết bạn với những người mà chúng ta có thể quen, nhằm mục đích tăng số lượng kết nối trên mạng xã hội. Kiểu gợi ý này với mục tiêu có chút khác biệt so với hệ thống gợi ý sản phẩm. Trong khi hệ thống gợi ý sản phẩm trực tiếp tạo ra lợi nhuận cho người cung cấp dịch vụ bằng cách tăng doanh số sản phẩm bán được, thì trong mạng xã hội, việc gợi ý kết bạn nhằm tăng số lượng kết nối trên mạng xã hội và từ đó nâng cao trải nghiệm của người dùng. Cùng với việc tăng được số lượng kết nối, lợi nhuận thu được từ việc quảng cáo cũng sẽ tăng lên.