

1. Knowledge based

1.1. Khái quát Knowledge

Knowledge-based recommendation (KBR) là một trong các phương pháp để đề xuất sản phẩm mới dựa trên các thông tin liên quan đến một sản phẩm nào đó.

Cụ thể hơn, KBR **không** sử dụng thông tin về đánh giá của người dùng để đưa ra đề xuất, mà thay vào đó sử dụng các thông tin mô tả sản phẩm, thông tin về đặc tính, tính năng của sản phẩm, và có thể suy rộng hơn là sử dụng thông tin từ những lĩnh vực khác, nhiều nguồn khác nhau có liên quan đến sản phẩm để tạo ra các đề xuất.

1.2. Phân tích Knowledge

Một số kỹ thuật đề xuất được sử dụng trong KBR bao gồm:

Thứ nhất, *Rule-based recommendation*: hay được biết đến là *Hệ gợi ý dựa trên quy tắc*, kỹ thuật này sử dụng các luật logic và các quy tắc để đưa ra các đề xuất sản phẩm phù hợp với yêu cầu của người dùng. Và các quy tắc này phải được xây dựng dựa trên các thuộc tính của sản phẩm (đúng như bản chất của Knowledge based).

Ví dụ: nếu người dùng muốn mua một chiếc xe đẹp để đi địa hình, thuật toán có thể đưa ra các đề xuất cho các loại xe đẹp có độ bền cao và chuyên dụng cho việc chạy trên những địa hình không bằng phẳng.

Thứ hai, *Case-based reasoning*: hay được biết đến là *Hệ lý luận dựa trên trường hợp*, thì đây là một quá trình lập luận và đưa ra quyết định một cách tự động, nhờ đó mà có thể giải quyết các đề xuất sản phẩm mới. Cụ thể hơn, bằng cách lưu trữ các dữ kiện và các hành vi của khách hàng được thực hiện tại thời điểm đó, nó thiết lập một tập hợp các quy tắc quy định riêng để thực hiện việc đề xuất. Kỹ thuật này được dùng trong việc giải quyết các bài toán phân loại, dự đoán, gợi ý, và khắc phục sự

cố. Một ví dụ về Case-based reasoning trong lĩnh vực SEO như sau: Giả sử một khách hàng đang mua sắm trực tuyến và tìm kiếm một đôi giày chạy bộ. Hệ thống gợi ý mà có sử dụng Case-based reasoning để giải quyết vấn đề này bằng cách tìm kiếm trong các hồ sơ đã lưu trữ những trường hợp tương tự với những khách hàng đã mua giày chạy bộ trước đó và có sở thích và nhu cầu tương tự để gợi ý đôi giày ưng ý với khách hàng hiện tại.

Thứ ba, *Semantic-based recommendation* (hệ để xuất dựa trên ngữ nghĩa): sử dụng các khái niệm và ý nghĩa của từng sản phẩm để đưa ra các đề xuất phù hợp với nhu cầu của người dùng. Thuật toán này phân tích nội dung của các sản phẩm và xử lý ngôn ngữ tự nhiên để hiểu ý nghĩa của từng sản phẩm, từ đó tìm ra các mối liên hệ liên quan đến nhu cầu và sở thích của người dùng.

(Ví dụ) Kỹ thuật này được áp dụng cho đề xuất bạn bè trên các trang mạng xã hội, sau khi phân tích về ngôn ngữ, quốc tịch, giọng nói,... thì gợi ý những bạn bè có các đặc điểm ngữ nghĩa tương tự.

1.3. Ưu điểm Knowledge

Một số ưu điểm của Knowledge-based recommendation đã được nhóm tìm hiểu bao gồm:

Thứ nhất, *Khả năng đưa ra các đề xuất sản phẩm chính xác và phù hợp với nhu cầu của người dùng*, bởi vì KBR sử dụng các thông tin kiến thức được train và thông tin từ nhiều nguồn khác nhau để đưa ra các đề xuất sản phẩm cho người dùng. Do đó sẽ không bị ảnh hưởng bởi những đánh giá giả mạo.

Thứ hai, *Không yêu cầu dữ liệu lớn để hoạt động hiệu quả*. So với Collaborative Filtering hay Content-based recommendation mà một lát nữa nhóm sẽ trình bày kỹ hơn, thì KBR không phụ thuộc vào số lượng dữ liệu người dùng và sản phẩm, do đó không bị ảnh hưởng bởi vấn đề lớn về dữ liệu thừa thãi.

Thứ ba, ***Khả năng giải quyết các vấn đề mới mà không cần dữ liệu lịch sử***, đây là một khả năng vô cùng tuyệt vời bởi vì KBR dựa trên kiến thức dữ liệu và thông tin về sản phẩm. Điều này rất hữu ích trong các lĩnh vực mới, sản phẩm mới hoặc với người dùng mới.

Thứ tư, ***Khả năng tùy chỉnh và cập nhật dữ liệu nhanh chóng***. KBR có thể được tùy chỉnh để đưa ra các đề xuất phù hợp với các nhu cầu cụ thể của người dùng. Ngoài ra, khi có thêm thông tin mới về sản phẩm, KBR có thể được cập nhật nhanh chóng để đưa ra các đề xuất mới.

1.4. Nhược điểm Knowledge

Vừa rồi nhóm vừa trình bày về các ưu điểm của **KBR**, tiếp theo hãy nói đến **nhược điểm** của kỹ thuật đề xuất này

Thứ nhất, ***Khả năng đưa ra các đề xuất đôi khi còn hạn chế***. Vì về bản chất KBR chỉ dựa trên thông tin kiến thức và thông tin sản phẩm để đưa ra các đề xuất, do đó nó có thể bỏ qua những thông tin quan trọng khác, như thông tin đánh giá từ người dùng.

Thứ hai, ***Bị giới hạn lại sự đa dạng trong các đề xuất***. Cũng nhờ vào tính chất của kỹ thuật này mà KBR chỉ dựa trên thông tin kiến thức và thông tin sản phẩm để đưa ra các đề xuất, do đó nó có thể bỏ qua các sản phẩm hoặc ý tưởng mới, chưa được biết đến trong ngành công nghiệp. Cũng như, KBR không thể đưa ra các đề xuất phù hợp với những người dùng mới vì nó không có thông tin về lịch sử đánh giá và sở thích của người dùng.

2. Hybrid

2.1. Khái quát Hybrid

Hybrid recommendation, hay được biết đến là hệ tư vấn lai ghép, là một phương pháp để xuất thông tin dựa trên sự kết hợp giữa các phương pháp khác nhau để đưa ra các kết quả để xuất tốt hơn. Thông thường, phương pháp này sẽ kết hợp hai hoặc nhiều loại phương pháp khác nhau, chẳng hạn như Collaborative Filtering và Content-Based Filtering.

Việc kết hợp các phương pháp khác nhau giúp tăng cường khả năng đưa ra các đề xuất chính xác và đa dạng hơn, vì mỗi phương pháp có những ưu điểm và nhược điểm riêng. Ví dụ, Collaborative Filtering có thể đưa ra các đề xuất phù hợp với sở thích của người dùng, trong khi Content Based Filtering có thể đưa ra các đề xuất phù hợp với đặc điểm của sản phẩm.

Chúng ta có thể lấy ví dụ từ TikTok

(tự phân tích slide ảnh chụp màn hình bài báo TikTok)

3. Content based

3.1. Khái niệm Content-based

Content-based filtering là một phương pháp lọc thông tin hay một hệ thống khuyến nghị dựa trên nội dung của các mục/sản phẩm đã được đánh giá.

Theo *The International Journal of Computer Applications - Volume 110, January 2015 - Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System* đã đề cập rằng “*The items are recommended based on the features of items to find the good match to other users...*”. Thì với mô hình content-based filtering, hệ thống khuyến nghị sử dụng thông tin về thuộc tính của các mục đã được đánh giá (ví dụ với các tiêu chí để đánh giá một bộ phim gồm có:

nội dung, thể loại, độ dài, tác giả, diễn viên, ...) để đề xuất các mục có nội dung tương tự cho người dùng.

Content-based filtering giả định rằng nếu một người dùng đã thích một mục cụ thể, thì họ sẽ thích các mục có nội dung tương tự trong tương lai. Do đó, hệ thống khuyến nghị sử dụng các thuật toán để tìm kiếm các mục có thuộc tính tương tự với mục mà người dùng đã đánh giá và đề xuất chúng cho người dùng.

3.2. Phân tích Content based

Công thức của Content-based filtering được sử dụng để tính toán độ tương đồng giữa các item, và có dạng như trên màn hình đang hiển thị:

$$\text{sim}(a,b) = \cos(\theta) = (a \cdot b) / (\|a\| \cdot \|b\|)$$

Trong đó:

$\text{sim}(a,b)$ là độ tương đồng giữa hai mặt hàng a và b.

$\cos(\theta)$ là cosine similarity giữa hai vector đặc trưng của a và b.

$a \cdot b$ là tích vô hướng giữa hai vector đặc trưng của a và b.

$\|a\|$ và $\|b\|$ là độ dài của hai vector đặc trưng của a và b.

Công thức trên tính toán độ tương quan cosine giữa hai vector đặc trưng của hai mặt hàng. Nó cho kết quả là một số thực trong khoảng từ -1 đến 1, với giá trị càng gần 1 thì hai mặt hàng càng giống nhau về mặt đặc trưng.

3.3. Ưu điểm Content based

Thứ nhất, *độc lập với người dùng*: Content-based filtering không yêu cầu thông tin về hành vi của người dùng để đưa ra các đề xuất. Điều này có nghĩa là nó

hoạt động tốt cho các người dùng mới và không yêu cầu người dùng cung cấp bất kỳ thông tin cá nhân nào. Ta thấy có điểm tương đồng giữa thuật toán này với Knowlegde based khi nãy.

Thứ hai, *khả năng tiên đoán tốt*: Vì content-based filtering tập trung vào đặc tính của một mục, nó có thể dự đoán tốt những mục mà người dùng có thể quan tâm. Ví dụ, nếu một người dùng thích một bộ phim nhất định, họ có thể thích các bộ phim khác có các đặc tính tương tự.

Thứ ba, *khả năng mở rộng*: Content-based filtering có thể mở rộng để bao gồm nhiều loại dữ liệu, bao gồm cả văn bản, âm nhạc và hình ảnh. Nó cũng có thể tự động tạo ra các đặc trưng của các mục dựa trên dữ liệu đầu vào, cho phép nó mở rộng đến các loại dữ liệu mới.

Thứ tư, *giải thích dễ dàng*: Content-based filtering cho phép giải thích dễ dàng cho người dùng vì nó dựa trên các đặc tính rõ ràng của các mục. Khi đề xuất được đưa ra, người dùng có thể dễ dàng thấy lý do tại sao các mục đó được đề xuất nhờ vào điểm số tương đồng (similarity score) đã được tính toán.

3.4. Nhược điểm Content based

Mặc dù Content-based Filtering là một phương pháp đánh giá và đề xuất sản phẩm rất hữu ích, nhưng nó cũng có một số nhược điểm như sau:

Thứ nhất, *Giới hạn của dữ liệu*: Content-based Filtering chỉ có thể dựa trên các thông tin về sản phẩm để đưa ra đề xuất, nhưng nó không thể tận dụng được các thông tin liên quan đến người dùng, ví dụ như thông tin về hành vi. Điều này có thể dẫn đến các đề xuất sản phẩm không chính xác hoặc không phù hợp.

Thứ hai, *Khó khăn trong việc tạo đặc trưng sản phẩm*: Để áp dụng Content-based Filtering, cần phải xác định các đặc trưng của sản phẩm, nhưng đôi khi việc tạo đặc trưng này khó khăn, đặc biệt là với những sản phẩm phức tạp và đa dạng, với những sản phẩm có nhiều tiêu chí để đánh giá.

Thứ ba, *Hạn chế của khả năng tìm kiếm*: Content-based Filtering chỉ có thể đề xuất sản phẩm tương tự với những sản phẩm đã được xem trước đó. Nó không thể đưa ra các đề xuất sản phẩm hoàn toàn mới hoặc các sản phẩm mà người dùng chưa từng xem trước đây

3.5. Ví dụ thực tế Content based

Một ví dụ thực tế cho phương pháp Content-based filtering là khi một người dùng đang mua sắm trực tuyến trên một trang web thương mại điện tử. Hệ thống có thể sử dụng thông tin về các sản phẩm mà người dùng đó đã xem hoặc mua để đề xuất các sản phẩm tương tự có cùng màu sắc, kiểu dáng, chất liệu hoặc thương hiệu. Hệ thống này sẽ dựa trên các đặc tính của sản phẩm để đưa ra các gợi ý sản phẩm tương tự mà người dùng này có thể quan tâm.

4. Collaborative Filtering

4.1. Khái quát Collaborative

Collaborative filtering là một kỹ thuật để xuất được sử dụng để giúp đưa ra các dự đoán hoặc đề xuất cho người dùng dựa trên sự tương tự, mức độ đánh giá giữa các người dùng hoặc sản phẩm.

Cũng theo *The International Journal of Computer Applications - Volume 110 – January 2015 - Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System* đã đề cập rằng “*The items are recommended based on rating of users who gave similar rating scores*”

Thì ý tưởng chính của collaborative filtering là xác định các mối liên hệ tương tự giữa các người dùng hoặc sản phẩm bằng cách phân tích các đánh giá hoặc sự tương tác giữa chúng. Dựa trên các mối liên hệ này, collaborative filtering sử dụng các thông tin đã được xác định trên hồ sơ những người dùng trước đó để đưa ra dự đoán hoặc đề xuất sản phẩm cho người dùng.

4.2. Phân tích Collaborative

Có hai phương pháp collaborative filtering chính:

- User-based collaborative filtering: dựa trên mức độ tương đồng (similarity score) giữa các người dùng để đưa ra các đề xuất cho người dùng mới dựa trên những gì người dùng khác đã đánh giá hay sử dụng trước đó.
- Item-based collaborative filtering: dựa trên mức độ tương đồng (similarity score) giữa các sản phẩm để đưa ra các đề xuất cho người dùng mới dựa trên các sản phẩm mà họ đã thích hoặc đã mua trước đó.

Trong *user-based collaborative filtering*, để dự đoán sở thích của một người dùng mới, ta tính toán trung bình trọng số các đánh giá của người dùng khác có sở thích tương tự với người dùng mới đó.

Công thức tính toán đánh giá dự đoán của một người dùng mới (u) đối với một sản phẩm (i) như sau:

$$r(u,i) = (\sum sim(u, v) * r(v, i)) / (\sum sim(u, v))$$

Trong đó:

$r(u,i)$ là đánh giá dự đoán của người dùng mới (u) đối với sản phẩm (i).

$sim(u, v)$ là độ tương đồng giữa người dùng mới (u) và người dùng khác (v).

$r(v, i)$ là đánh giá của người dùng khác (v) đối với sản phẩm (i).

Còn trong *item-based collaborative filtering*, để đưa ra đề xuất cho một người dùng mới, ta tính toán trung bình trọng số các đánh giá của người dùng đó đối với các sản phẩm có tính chất tương tự với sản phẩm được đề xuất.

Công thức tính toán đánh giá dự đoán của một người dùng (u) đối với một sản phẩm (i) như sau:

$$r(u,i) = (\sum sim(i,j) * r(u,j)) / (\sum sim(i,j))$$

Trong đó:

$r(u,i)$ là đánh giá dự đoán của người dùng (u) đối với sản phẩm (i).

$sim(i,j)$ là độ tương đồng giữa sản phẩm (i) và sản phẩm khác (j).

$r(u,j)$ là đánh giá của người dùng (u) đối với sản phẩm khác (j).

4.3. Ưu điểm Collaborative

Thứ nhất, *Khả năng đề xuất sản phẩm tốt*: Collaborative filtering dựa trên thông tin về sự tương tác giữa người dùng hoặc sản phẩm để đưa ra các đề xuất cho người dùng. Điều này giúp đề xuất sản phẩm chính xác và đáng tin cậy cho người dùng, giúp họ tìm kiếm các sản phẩm phù hợp với sở thích của mình.

Thứ hai, *Khả năng mở rộng*: Collaborative filtering có thể được áp dụng cho các tập dữ liệu lớn và phức tạp, giúp nó trở thành một trong những phương pháp được sử dụng nhiều nhất để đề xuất sản phẩm.

Thứ ba, *Khả năng hoạt động độc lập với nội dung sản phẩm*: Collaborative filtering không cần phải biết thông tin về nội dung của sản phẩm, điều này giúp nó có thể hoạt động hiệu quả với các loại sản phẩm khác nhau.

Thứ tư, *Khả năng thích nghi với sở thích của người dùng*: Collaborative filtering có thể thích nghi với sở thích của người dùng bằng cách đưa ra các đề xuất dựa trên lịch sử tương tác của người dùng

4.4. Nhược điểm Collaborative

Mặc dù kỹ thuật Collaborative Filtering có nhiều ưu điểm tích cực, nhưng nó cũng có một số nhược điểm dẫn đến các sai lệch trong kết quả đề xuất:

Thứ nhất, *Vấn đề khởi động lạnh (cold-start problem)*: Khi một sản phẩm mới được giới thiệu hoặc một người dùng mới tham gia, Collaborative Filtering không có đủ thông tin để đưa ra đề xuất sản phẩm. Điều này gọi là vấn đề khởi động lạnh.

Thứ hai, *Vấn đề xếp hạng và đánh giá*: Việc đánh giá và xếp hạng của người dùng có thể không chính xác hoặc bị ảnh hưởng bởi nhiều yếu tố. Ví dụ, một người dùng có thể đánh giá sản phẩm xấu chỉ vì họ không thích một thuộc tính nhất định của sản phẩm.

Thứ ba, *Vấn đề lượng dữ liệu*: Để đưa ra đề xuất chính xác, CF yêu cầu một lượng dữ liệu đủ lớn. Nếu lượng dữ liệu quá ít, hệ thống đề xuất sẽ không hoạt động tốt.

Thứ tư, *Vấn đề giả mạo và spam*: CF có thể bị ảnh hưởng bởi các đánh giá giả mạo hoặc spam. Nếu một người dùng tạo nhiều tài khoản giả để đánh giá sản phẩm của mình, hệ thống đề xuất sẽ không hoạt động tốt.

4.5. Ví dụ thực tế:

Ta có một ví dụ về đề xuất lọc cộng tác dựa trên người dùng, rằng giả sử có một người dùng A thường xem các bộ phim hành động và một người dùng B thường xem các bộ phim kinh dị, nhưng cả hai đều thích bộ phim "The Dark Knight", hệ thống sẽ đề xuất cho người dùng A các bộ phim khác mà người dùng B thường xem và thích. Đây là một ví dụ về collaborative filtering trong lĩnh vực phim ảnh.

4.6. Hiệu quả của phương pháp lai ghép giữa Content Based Filtering với Collaborative Filtering

Việc lai ghép, kết hợp giữa hai phương pháp Content-Based Filtering (CBF) với Collaborative Filtering (CF) Recommendation kề trên có thể đem lại nhiều lợi ích, bao gồm:

Thứ nhất, *Giúp đa dạng hóa các đề xuất*: Khi kết hợp CBF và CF, ta có thể đưa ra các đề xuất đa dạng và phù hợp với sở thích của người dùng. CBF có thể đưa ra các đề xuất dựa trên các đặc trưng của sản phẩm, trong khi CF có thể đưa ra các đề xuất dựa trên sự tương tự giữa các người dùng. Kết hợp cả hai phương pháp cho phép đưa ra các đề xuất phù hợp với sở thích và nhu cầu của người dùng.

Ngoài ra việc kết hợp còn, *Cải thiện độ chính xác của đề xuất*: CF có thể đưa ra các đề xuất dựa trên hành vi của những người dùng có sở thích tương tự, trong khi CBF có thể đưa ra các đề xuất dựa trên nội dung của sản phẩm. Khi được kết hợp lại, ta có thể sử dụng các đặc trưng của sản phẩm và thông tin về hành vi của người dùng để cải thiện độ chính xác của các đề xuất.

Và cuối cùng là, *Tăng tính linh hoạt của hệ thống*: Khi kết hợp CBF và CF, ta có thể tăng tính linh hoạt của hệ thống đề xuất. Ví dụ, nếu một sản phẩm mới được phát hành và không có đủ dữ liệu về hành vi của người dùng, ta vẫn có thể đưa ra các đề xuất phù hợp bằng cách sử dụng các đặc trưng của sản phẩm. Hay nôm na là bù qua sót lại.

4.7. Điểm khác biệt giữa Content-based recommendation với Knowledge-based recommendation

Nhóm chúng em còn phân tích một chút về điểm khác biệt giữa Knowledge-based recommendation (KBR) và Content-based recommendation (CBR). Thì mặc

dù cả hai phương pháp này đều dựa trên thông tin về sản phẩm, nhưng cách tiếp cận và thuật toán sử dụng để đưa ra các đề xuất khác nhau. Cụ thể đã được tập thể nhóm Neko tìm hiểu như sau:

Content-based recommendation (CBR) sử dụng các thông tin về sản phẩm như đặc tính, tính năng, và thông tin mô tả để tạo ra các đề xuất. CBR tập trung vào sự tương đồng giữa các sản phẩm và sở thích của người dùng. Nó tìm kiếm các sản phẩm có các đặc tính giống nhau với các sản phẩm mà người dùng đã thích trước đó để đưa ra các đề xuất mới.

Trong khi đó, KBR không chỉ tập trung vào các đặc tính của sản phẩm, mà còn sử dụng các thông tin kiến thức để tạo ra các đề xuất sản phẩm. KBR sử dụng thông tin từ các nguồn tri thức khác để xây dựng một mô hình tri thức về các sản phẩm và người dùng. KBR dựa trên thông tin kiến thức để đưa ra các đề xuất sản phẩm có thể phù hợp với nhu cầu của người dùng.

Tóm lại, KBR và CBR đều sử dụng thông tin sản phẩm để đưa ra các đề xuất, tuy nhiên KBR còn sử dụng thêm các thông tin kiến thức từ nhiều nguồn được đưa vào để xác định các sản phẩm phù hợp.

Lấy ví dụ Spotify

...

Demo 2 source code php

(trong 2 folder **htdocs: /demo_recommendation_system** và **/content_based**)

(Còn lại tự nói)