\***Cách triển khai thuật toán**

**Collaborative filtering based on neighbor.**

**Hướng user-based**

Offline: Ma trận đánh giá user-item

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | i1 | i2 | i3 | i4 | i5 |
| u1 | 1 | 2 | \* | \* | 3 |
| u2 | 2 | 3 | \* | 1 | 1 |
| u3 | \* | 3 | 2 | \* | 4 |
| u4 | 2 | \* | 5 | 2 | 1 |
| u5 | 2 | \* | 4 | 5 | 3 |

-Tính độ tương đồng giữa những người dùng với nhau

Sim(ui,uj) với mọi i,j

-Với mỗi người dùng, chúng ta xếp hạng các người dùng tương đồng với anh ấy dựa trên SIM

Online: Target user u1

Tính ru1,i3

Dựa vào dữ liệu đã có ở offline, chúng ta tìm được k người dùng tương đồng với u1 mà có đánh giá i3.

* Tổng hợp đánh giá

Tính ru1,i4

Dựa vào dữ liệu đã có ở offline, chúng ta tìm được k người dùng tương đồng với u1 mà có đánh giá i4.

* Tổng hợp đánh giá

**Cách triển khai thuật toán**

**Collaborative filtering based on neighbor.**

**Hướng item-based**

Offline: Ma trận đánh giá user-item

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | i1 | i2 | i3 | i4 | i5 |
| u1 | 1 | 2 | \* | \* | 3 |
| u2 | 2 | 3 | \* | 1 | 1 |
| u3 | \* | 3 | 2 | \* | 4 |
| u4 | 2 | \* | \* | 2 | 1 |
| u5 | 2 | \* | \* | 5 | 3 |

-Tính độ tương đồng giữa những sản phẩm với nhau

Sim(ia,ib) với mọi a,b

-Với mỗi sản phẩm, chúng ta xếp hạng các sản phẩm với nó dựa trên SIM

Online: Target user u1

Tính ru1,i3

Dựa vào dữ liệu đã có ở offline, chúng ta tìm được k sản phẩm tương đồng với i3 mà u1 có đánh giá.

* Tổng hợp đánh giá

Tính ru1,i4

Dựa vào dữ liệu đã có ở offline, chúng ta tìm được k sản phẩm tương đồng với i4 mà có đánh giá u1.

* Tổng hợp đánh giá

So sánh **Collaborative filtering based on neighbor**

**hướng user-based và Collaborative filtering based on neighbor hướng item-based.**

-Chi phí tính toán: số lượng item sẽ ít hơn số lượng user 🡺 hướng item-based tiết kiệm hơn

-Độ đa đạng của danh sách tư vấn: hướng user-based đa đang hướng item-based 🡪 ít nhàm chán.

-Tình diễn giải kết quả tư vấn:

-item-based: do bạn thích sản phẩm A, sản phẩm B nên tôi dự đoán được bạn thích sản phẩm C 🡺 dễ dàng diễn giải.

-User-based: do người dùng X thích A, người Y thích A, mà X,Y có vị tương đồng với bạn nên tôi tư vấn cho bạn sản phẩm A 🡺 khó diễn giải.

-Tính ổn định của kết quả offiline: user-based << item-based.

Đánh giá điểm mạnh và điểm yếu của **Collaborative filtering based on neighbor**

Điểm mạnh: đơn giản và gần với con người.

Điểm yếu: -khó triển khai trong hệ thống lớn (tốt 🡪 đầu tư nhiều về mặt phần cứng)

* Tính thưa của ma trận user-item 🡪 khó tính độ tương đồng
* Tính thưa của ma trận user-item 🡪 khó tổng hợp đánh giá
* Sản phẩm mới --> không chạy được quá trình tư vấn cho sản phẩm đó.
* Người dùng mới 🡪 không chạy được quá trình tư vấn cho người dùng đó.

Tính độ tương đồng.

Công thức pearson nên được điều chỉnh dựa trên độ tin cậy (số lượng đánh giá chung của 2 người dùng)

Sim(u,v)= reliability(u,v). pearson(u,v).

reliability(u,v) chạy từ (0,1]

reliability(u,v)=min( số lượng sản phẩm đánh giá chung của u và v, Beta)/Beta.