

Trong đó:

\hat{r}_{ui} chính là dự đoán cho người dùng u trên sản phẩm i

$sim(u, u')$: độ tương tự giữa người dùng u và u' .

K_u là số người dùng có độ lân cận gần người dùng u .

Chúng tôi biểu diễn giải thuật lọc cộng tác dựa trên người dùng lân cận (User_KNN) sử dụng độ tương tự Pearson bằng ngôn ngữ giả để dự đoán độ thích cho người dùng u trên sản phẩm i như sau:

```

1: procedure USERKNN-CF( $\bar{r}_u, r, D^{train}$ )
2:   for  $u=1$  to  $N$  do
3:     Tính  $Sim_{uu'}$ , sử dụng công thức (2)
4:   end for
5:   Sort  $Sim_{uu'}$  // sắp xếp giảm dần độ tương tự
6:   for  $k=1$  to  $K$  do
7:      $K_u \leftarrow k$  // Các người dùng  $k$  gần nhất của  $u$ 
8:   end for
9:   for  $i = 1$  to  $M$  do
10:    Tính  $\hat{r}_{ui}$ , sử dụng công thức (4)
11:   end for
12: end procedure

```

Trong đó:

\bar{r}_u : đánh giá trung bình của người dùng u trên tất cả các item

r : đánh giá của người dùng trên tập huấn luyện

K : người dùng k gần nhất

N : số người dùng

M : số item

D^{train} : tập dữ liệu huấn luyện

4 PHƯƠNG PHÁP SỬ DỤNG LUẬT BÌNH CHỌN SỐ ĐỒNG KẾT HỢP NGƯỖNG SỞ THÍCH

Trong phần này chúng tôi sẽ đề xuất phương pháp Luật bình chọn số đồng kết hợp ngưỡng sở thích (đặt tên là MASSVOTING) dựa trên ý tưởng xem xét sự tương đồng về sở thích của các user đối với các item, được mô tả cụ thể như sau:

– Giả sử, cho ma trận dữ liệu user-item-rating như Hình 2.

– Gọi U là tập tất cả các user có trong cơ sở dữ liệu.

– Xét tại user u

+ Gọi I_u là tập các item cần dự đoán xếp hạng của user u và $I_{u'}$ là tập các item của user u' tương ứng với các item cần dự đoán xếp hạng của user u có giá trị xếp hạng $\geq T$ (với T là ngưỡng sở thích nhận giá trị thuộc tập $R = \{1, 2, 3, 4, 5\}$).

+ Gọi $I_{uu'}$ là tập các item được đánh giá bởi cả user u và u' (với $u' \in U$) có sở thích $\geq T$.

+ Xét item i thuộc $I_{u'}$, với mỗi item i' thuộc $I_{uu'}$ tăng giá trị (số lần được đánh giá xếp hạng $\geq T$) cho item thứ i trong tập I_u lên 1.

Sau khi duyệt qua tất cả các user, mỗi item trong tập I_u sẽ chứa số lần được đánh giá xếp hạng $\geq T$.

Sắp xếp I_u theo thứ tự giảm dần và chọn ra top N items để gợi ý cho user u .

Item/ User	1	2	3	4	5
1	1	4	5	?	3
2	5	1	?	5	2
3	4	1	2	5	?
4	?	3	4	?	4
5	2	1	?	5	2
6	?	4	1	2	4
7	5	3	2	1	?
8	3	4	4	?	5
9	2	4	4	2	?
10	3	4	1	?	2

Hình 2: Ma trận dữ liệu user-item-rating

Giải thuật luật bình chọn số đồng kết hợp ngưỡng sở thích để gợi ý item i cho user u được biểu diễn bằng ngôn ngữ giả như sau:

```

1: function MASSVOTING( $D^{train}, u, i, threshold$ )
2:    $\hat{r}_{ui} = 0$  // Khởi tạo giá trị cần dự đoán
3:   for  $u'=1$  to  $N$  do // Duyệt qua tất cả user ngoại trừ user  $u$ 
4:     if ( $r_{u'i} \geq threshold$ ) then // Nếu giá trị xếp hạng của user  $u'$  cho item  $i$  lớn hơn hoặc bằng ngưỡng
5:        $\hat{r}_{ui} \leftarrow count(I_{uu'})$  // Đếm số item trong  $I_{uu'}$  và cộng dồn vào  $\hat{r}_{ui}$ 
6:     end if
7:   end for

```