

## Worked out example: Item based Collaborative filtering for Recommender Engine

Item based collaborative filtering is a model-based algorithm for recommender engines. In item based collaborative filtering similarities between items are calculated from rating-matrix. And based upon these similarities, user's preference for an item not rated by him is calculated. Here is a step-by-step worked out example for four users and three items. We will consider the following sample data of preference of four users for three items:

ID	user	item	rating
241	u1	m1	2
222	u1	m3	3
276	u2	m1	5
273	u2	m2	2
200	u3	m1	3
229	u3	m2	3
231	u3	m3	1
239	u4	m2	2
286	u4	m3	2

**Step 1:** Write the user-item ratings data in a matrix form. The above table gets rewritten as follows:

	m1	m2	m3
u1	2	?	3
u2	5	2	?
u3	3	3	1
u4	?	2	2

Here rating of user u1 for item m3 is 3. There is no rating for item m2 by user u1. And no rating also for item m3 by user u2.

**Step 2:** We will now create an item-to-item similarity matrix. The idea is to calculate how similar an item is to another item. There are a number of ways of calculating this. We will use [cosine similarity measure](#). To calculate similarity between items m1 and m2, for example, look at all those users who have rated both these items. In our case, both m1 and m2 have been rated by users u2 and u3. We create two item-vectors, v1 for item m1 and v2 for item m2, in the user-space of (u2,u3) and then find the cosine of angle between these vectors. A zero angle or overlapping vectors with cosine value of 1 means total similarity (or per user, across all items, there is same rating) and an angle of 90 degree would mean cosine of 0 or no similarity. Thus, the two item-vectors would be,

$$\begin{aligned}v_1 &= 5 u_2 + 3 u_3 \\v_2 &= 3 u_2 + 3 u_3\end{aligned}$$

The cosine similarity between the two vectors,  $v_1$  and  $v_2$ , would then be:

$$\cos(v_1, v_2) = (5*3 + 3*3) / \sqrt{(25 + 9) * (9+9)} = 0.76$$

Similarly, to calculate similarity between  $m_1$  and  $m_3$ , we consider only users  $u_1$  and  $u_3$  who have rated both these items. The two item vectors,  $v_1$  for item  $m_1$  and  $v_3$  for item  $m_3$ , in the user-space would be as follows:

$$\begin{aligned}v_1 &= 2 u_1 + 3 u_3 \\v_3 &= 3 u_1 + 1 u_3\end{aligned}$$

The cosine similarity measure between  $v_1$  and  $v_3$  is:

$$\cos(v_1, v_3) = (2*3 + 3*1) / \sqrt{(4 + 9) * (9+1)} = 0.78$$

We can similarly calculate similarity between items  $m_2$  and  $m_3$  using ratings given to both by users  $u_3$  and  $u_4$ . The two item-vectors  $v_2$  and  $v_3$  would be:

$$\begin{aligned}v_2 &= 3 u_3 + 2 u_4 \\v_3 &= 1 u_3 + 2 u_4\end{aligned}$$

And cosine similarity between them is:

$$\cos(v_2, v_3) = (3*1 + 2*2) / \sqrt{(9 + 4) * (1 + 4)} = 0.86$$

We now have the complete item-to-item similarity matrix as follows:

	m1	m2	m3
m1	1	0.76	0.78
m2	0.76	1	0.86
m3	0.78	0.86	1

**Step 3:** For each user, we next predict his ratings for items that he had not rated. We will calculate rating for user  $u_1$  in the case of item  $m_2$  (target item). To calculate this we weigh the just-calculated similarity-measure between the target item and other items that user has already rated. The weighing factor is the ratings given by the user to items already rated by him. We further scale this weighted sum with the sum of similarity-measures so that the calculated rating remains within a predefined limits. Thus, the predicted rating for item  $m_2$  for user  $u_1$  would be calculated using similarity measures between  $(m_2, m_1)$  and  $(m_2, m_3)$  weighted by the respective ratings for  $m_1$  and  $m_3$ :

$$\text{rating} = (2 * 0.76 + 3 * 0.86) / (0.76 + 0.86) = 2.53$$