ITEM BASED FILTERING SYSTEM

BY:

M. Ashish Reddy

DSWP -> Batch-5

Collaborative Filtering is a technique or a method to predict a user's taste and find the items that a user might prefer on the basis of information collected from various other users having similar tastes or preferences. It takes into consideration the basic fact that if person X and person Y have a certain reaction for some items then they might have the same opinion for other items too.

The two most popular forms of collaborative filtering are:

- **User Based:** Here, we look for the users who have rated various items in the same way and then find the rating of the missing item with the help of these users.
- **Item Based:** Here, we explore the relationship between the pair of items (the user who bought Y, also bought Z). We find the missing rating with the help of the ratings given to the other items by the user.

Let's talk about Item-Based Collaborative Filtering in detail. It was first invented and used by Amazon in 1998. Rather than matching the user to similar customers, item-to-item collaborative filtering matches each of the user's purchased and rated items to similar items, then combines those similar items into a recommendation list. Now, let us discuss how it works.

Item to Item Similarity: The very first step is to build the model by finding similarity between all the item pairs. The similarity between item pairs can be found in different ways. One of the most common methods is to use cosine similarity.

Formula for Cosine Similarity:

Prediction Computation: The second stage involves executing a recommendation system. It uses the items (already rated by the user) that are most similar to the missing item to generate rating. We hence try to generate predictions based on the ratings of similar products. We compute this using a formula which computes rating for a particular item using weighted sum of the ratings of the other similar products.

Example:
Let us consider one example. Given below is a set table that contains some items and the user who have rated those items. The rating is explicit and is on a scale of 1 to 5. Each entry in the table denotes the rating given by a ith User to a jth Item. In most cases majority of cells are empty as a user rates only for few items. Here, we have taken 4 users and 3 items. We need to find the missing ratings for the respective user.

User/Item	Item_1	Item_2	Item_3
User_1	2	–	3
User_2	5	2	–
User_3	3	3	1
User_4	–	2	2

Step 1: Finding similarities of all the item pairs.
Form the item pairs. For example in this example the item pairs are (Item_1, Item_2), (Item_1, Item_3), and (Item_2, Item_3). Select each item to pair one by one. After this, we find all the users who have rated for both the items in the item pair. Form a vector for each item and calculate the similarity between the two items using the cosine formula stated above.

```
Sim(Item1, Item2)
```

In the table, we can see only User_1 and User_2 have rated for both items 1 and 2.

Thus, let I1 be vector for Item_1 and I2 be for Item_2. Then,

I1 = 5U2 + 3U3 and,

I2 = 2U2 + 3U3

$$Similarity(I1, I2) = \frac{(5*2)+(3*3)}{\sqrt{5^2+3^2}\sqrt{2^2+3^2}} = 0.90$$

Sim(Item2, Item3)

In the table we can see only User_3 and User_4 have rated for both the items 1 and 2.

Thus, let I2 be vector for Item_2 and I3 be for Item_3. Then,

I2 = 3U3 + 2U4 and,

I3 = 1U3 + 2U4

$$Similarity(I2, I3) = \frac{(3*1)+(2*2)}{\sqrt{3^2+2^2}\sqrt{1^2+2^2}} = 0.869$$

Sim(Item1, Item3)

In the table we can see only User_1 and User_3 have rated for both the items 1 and 2.

Thus, let I1 be vector for Item_1 and I3 be for Item_3. Then,

I1 = 2U1 + 3U3 and,

I3 = 3U1 + 1U3

$$Similarity(I1, I3) = \frac{(2*3)+(3*1)}{\sqrt{2^2+3^2}\sqrt{3^2+1^2}} = 0.789$$

Step 2: Generating the missing ratings in the table
Now, in this step we calculate the ratings that are missing in the table.

Rating of Item_2 for User_1

$$r(U_1, I_2) = \frac{r(U_1,I_1)*s_{I_1 I_2}+r(U_1,I_3)*s_{I_3 I_2}}{s_{I_1 I_2}+s_{I_3 I_2}} = \frac{(2*0.9)+(3*0.869)}{(0.9+0.869)} = 2.49$$

Rating of Item_3 for User_2

$$r(U_2, I_3) = \frac{r(U_2,I_1)*s_{I_1 I_3}+r(U_2,I_2)*s_{I_2 I_3}}{s_{I_1 I_3}+s_{I_2 I_3}} = \frac{(5*0.789)+(2*0.869)}{(0.789+0.869)} = 3.43$$

Rating of Item_1 for User_4

$$r(U_4, I_1) = \frac{r(U_4,I_2)*s_{I_1 I_2}+r(U_4.I_3)*s_{I_1 I_3}}{s_{I_1 I_2}+s_{I_1 I_3}} = \frac{(2*0.9)+(2*0.789)}{(0.9+0.789)} = 2.0$$

Conclusion

A recommender system or recommendation system is a subclass of information filtering systems that predict the items the user may be interested in based on the user past behaviour.

Collaborative filtering is one such recommendation technique that filters items of user interest based on user/item similarity. Due to ease of use and domain-free, it is being used and explored at a large scale by researchers.

In this blog, we have implemented item-based collaborative filtering to recommend movies to users using cosine similarity. Other similarity metrics such as the Pearson correlation coefficient and Jaccard similarity could also be explored. This is still an open area of research with the motive to provide the user with the most relevant items.