Tutorial Business Analytics

Homework 11 – Solution

Exercise 11.4 - Collaborative Filtering

The following ratings of the users 1, 2, 3, and the active user 4 are given for the products A to J. The ratings can assume values in {1, 2, ..., 10}.

Product	User			Active User
	1	**2**	**3**	**4**
A	5	7	9	10
B	X	9	7	10
C	1	10	2	X
D	8	8	9	9
E	X	X	5	1
F	7	X	X	3
G	9	9	8	X
H	X	7	8	6
I	10	2	8	1
J	1	4	6	6

x: no rating available

For the significance weight of user a and i, $s_{a,i}$, and the number of co-rated items, m, we set:

$$s_{a,i} = 1, \quad \text{if } m \geq 7$$
$$s_{a,i} = \frac{m}{7}, \quad \text{else}$$

a) Determine the **weighted correlations**, $w_{i,4}$, between the users $i \in \{1,2,3\}$ and the active user $a = 4$.

b) Determine the **rating predictions**, $p_a(\text{product})$, for product C and product G, taking into account the two users most similar to the active user.

c) What problems are associated with the application of collaborative filtering in practice? Name at least three.

d) What alternative recommending technique is there where these problems do not occur?

Solution:

a) The weighted correlation for users 1 and 4 is given by
$$w_{1,4} = s_{1,4} \cdot c_{1,4},$$

where the set of co-rated items is $I_{1,4} = \{A, D, F, I, J\}$. Let r_1 and r_4 be the ratings of users 1 and 4 for $I_{1,4}$, respectively. Then we get
$$w_{1,4} = \frac{5}{7} \cdot \frac{\operatorname{cov}(r_1, r_4)}{\sigma_{r_1} \cdot \sigma_{r_4}}$$
$$\approx -0.2832.$$

Analogously, we get $I_{2,4} = \{A, B, D, H, I, J\}$, $I_{3,4} = \{A, B, D, E, H, I, J\}$, and therefore $s_{2,4} = 6/7$ and $s_{3,4} = 1$. Thus,
$$w_{2,4} \approx 0.7687,$$
$$w_{3,4} \approx 0.5261.$$

Note: Due to symmetry we have $w_{i,j} = w_{j,i}$.

b) Here, we denote $r_i(C)$ as the rating of user i for product C and $\bar{r}_i$ as the average rating of user i.

The two most similar users to the active user are users 2 and 3. The rating forecast for product C of user 4, based on these users, then is
$$p_4(C) = \bar{r}_4 + \frac{\sum_{i=2}^{3} w_{4,u}(r_i(C) - \bar{r}_i)}{\sum_{i=2}^{3} |w_{4,i}|}$$
$$\approx 5.75 + \frac{0.7687(10 - 7) + 0.5261(2 - 6.8889)}{1.2948}$$
$$\approx 5.5446.$$

Analogously, for product G we get
$$p_4(G) \approx 5.75 + \frac{0.7687(9 - 7) + 0.5261(8 - 6.8889)}{1.2948}$$
$$\approx 7.3888.$$

c) Problems with the application of collaborative filtering
   i. **Cold Start:** Enough users and ratings are needed to generate recommendations.
   ii. **Sparsity:** Even with a very large number of users one has the problem that with a very large number of possible products the product-user-matrix is sparsely populated, so that it is difficult to find users who have rated the same products.
   iii. **First Rater:** With a new product, there must first be consumers who test and then evaluate it.
   iv. **Popularity Bias:** Users with unique tastes are not considered by the existing similarity measure. Only popular products are recommended.

d) Alternative recommending technique: Content-based recommending.