mining massive dataset - 3

Martha Lee

November 2024

1 exercise3

(a) The Pearson correlation similarity between user C and another user X is given by:

$$\text{sim}(C, X) = \frac{\sum_{i \in I}(r_{Ci} - \bar{r}_C)(r_{Xi} - \bar{r}_X)}{\sqrt{\sum_{i \in I}(r_{Ci} - \bar{r}_C)^2}\sqrt{\sum_{i \in I}(r_{Xi} - \bar{r}_X)^2}}$$

So that, Similarity between C and A is -0.666 Similarity between C and B is -0.725 Similarity between C and D is 0.851 The top-2 most similar users are D and A. Then, we can obtain,

$$r_{Ce} = 3 + \frac{0.851 \times (4-3) + (-0.666) \times (1-3)}{0.851 + |-0.666|}$$

$$\approx 4.84$$

(b) We calculate the similarity between item e and each item rated by user C. After calculating, we can find that the top-2 items are f and h with similarities 0.500 and 0.189. Then, we can obtain,

$$r_{Ce} = 3 + \frac{0.500 \times (5 - 3.25) + 0.189 \times (3-2)}{0.500 + 0.189}$$

$$\approx 3.88$$

The rating predicted by User-User Collaborative Filtering method is 4.84 for item e by user C, and the rating predicted by Item-Item Collaborative Filtering method is 3.88.