

UNIVERSITEIT GENT

RECOMMENDER SYSTEMS

# Assignment 1 - Non Personalized Rec

*Tibo Vanheule*

*2022-2023*

- 1 Consider the movie with ID 1. What is the value of the simple product association for the movie with ID 1064?**

Simple product association (1,1064): 0.08281174771304767

- 2 Consider the movie with ID 1. What is the value of the advanced product association for the movie with ID 1064?**

Advanced product association (1,1064): 8.869809626670484

- 3 Explain the difference between the values of the simple and advanced product association (question 1 and 2). How do we have to interpret these numbers?**

### **3.1 Simple association**

Simple association is an asymmetrical relationship that expresses the percentage of people who bought X, who also bought Y. X is something that occurred, what is now the probability that Y occurs?

Products with the highest percentage are recommended.

### **3.2 Advanced association**

Advanced product association calculates if X makes Y more likely than other products.

To interpret advanced association we follow these rules:

1.  $\text{Advassociation} > 1 \rightarrow Y$  is likely to be bought if item X is bought.
2.  $\text{Advassociation} = 1 \rightarrow Y$  no association between items.
3.  $\text{Advassociation} < 1 \rightarrow Y$  is unlikely to be bought if item X is bought.

- 4 Consider the movie with ID 1. What is the value of the simple product association for the movie with ID 2858?**

Simple product association (1,2858): 0.6783822821376986

**5 What are the movie titles and genres of movies with ID 1, 1064, and 2858?**

| movie ID | title                                  | genre                       |
|----------|--|-----------------------------|
| 1        | Toy Story (1995)                       | Animation Children's Comedy |
| 1064     | Aladdin and the King of Thieves (1996) | Animation Children's Comedy |
| 2858     | American Beauty (1999)                 | Comedy Drama                |

**6 Compare the results of question 1 and 4. Which movie has the highest simple association value (1064 or 2858) with movie ID 1 and explain why?**

Simple product association (1,1064): 0.08281174771304767  
Simple product association (1,2858): 0.6783822821376986

The movie American Beauty (1999) with an id of 2858, has a higher association. 67% who have rated X has also rated Y. But looking at the genre, the movies are fundamentally different. Probably we see the popularity problem, where American beauty is much more popular. Y has been rated because it is popular and not because X has been rated.

**7 Consider the movie with ID 1. What is the value of the advanced product association for the movie with ID 2858?**

Advanced product association (1,2858): 1.331564628089004

**8 Compare the results of question 2 and 7. Which movie has the highest advanced association value (1064 or 2858) with movie ID 1 and explain why?**

Advanced product association (1,1064): 8.869809626670484  
Advanced product association (1,2858): 1.331564628089004

Aladdin and the King of Thieves (1064) has a higher association. American beauty (2858) has a much lower association. It is closer to 1 indicating a low product association.

Alladin is more likely to be bought together with Toy story.

- 9 Calculate the top 10 most frequently rated movies. Provide the movie ID, number of users who rated the movie, and title for each.

| movie ID | frequency | title   | genre                               |
|----------|-----------|---|-------------------------------------|
| 2858     | 3428      | American Beauty (1999)                            | Comedy Drama                        |
| 260      | 2991      | Star Wars: Episode IV - A New Hope (1977)         | Action Adventure Fantasy Sci-Fi     |
| 1196     | 2990      | Star Wars: Episode V - The Empire Strikes Back... | Action Adventure Drama Sci-Fi War   |
| 1210     | 2883      | Star Wars: Episode VI - Return of the Jedi (1983) | Action Adventure Romance Sci-Fi War |
| 480      | 2672      | Jurassic Park (1993)                              | Action Adventure Sci-Fi             |
| 2028     | 2653      | Saving Private Ryan (1998)                        | Action Drama War                    |
| 589      | 2649      | Terminator 2: Judgment Day (1991)                 | Action Sci-Fi Thriller              |
| 2571     | 2590      | Matrix, The (1999)                                | Action Sci-Fi Thriller              |
| 1270     | 2583      | Back to the Future (1985)                         | Comedy Sci-Fi                       |
| 593      | 2578      | Silence of the Lambs, The (1991)                  | Drama Thriller                      |

- 10 Now you have to find the movies that can be associated to the movie with id 3941. Calculate the Top 5 movies with the highest simple association value. Provide the movie ID, association value, and title for each.

| movie ID | simple   | title                                     | genre                            |
|----------|----------|---|----------------------------------|
| 110      | 0.954545 | Braveheart (1995)                         | Action Drama War                 |
| 480      | 0.909091 | Jurassic Park (1993)                      | Action Adventure Sci-Fi          |
| 589      | 0.863636 | Terminator 2: Judgment Day (1991)         | Action Sci-Fi Thriller           |
| 1544     | 0.818182 | Lost World: Jurassic Park, The (1997)     | Action Adventure Sci-Fi Thriller |
| 260      | 0.818182 | Star Wars: Episode IV - A New Hope (1977) | Action Adventure Fantasy Sci-Fi  |

- 11 For the same movie as in question 10 (id 3941), calculate the Top 5 movies with the highest advanced association value. Provide the movie ID, association value, and title for each.

| movie ID | advanced   | title                                  | genre         |
|----------|------------|--|---------------|
| 3940     | 547.090909 | Slumber Party Massacre III, The (1990) | Horror        |
| 3938     | 465.027273 | Slumber Party Massacre, The (1982)     | Horror        |
| 3939     | 437.672727 | Slumber Party Massacre II, The (1987)  | Horror        |
| 3942     | 422.752066 | Sorority House Massacre II (1990)      | Horror        |
| 3777     | 156.311688 | Nekromantik (1987)                     | Comedy Horror |

**12 Compare the resulting lists of question 9 and 10. What do you witness and how can you explain this?**

Many of the movies in the table of question 10 are popular movies (most rated) which are listed in the table of question 9. This is a short coming of the simple product association. The problem is that Y is rated regardless of X because they are popular not that X makes Y more likely.

**13 Compare the resulting lists of question 10 and 11. Which one is the best (=most accurate) according to you, and why?**

Since the table of 11 expresses how likely Y is, regardless of popularity, makes this table more accurate. The movies also share the same genre (horror). More note worthy the successor of "Sorority House Massacre" (3941) is included in the list of question 11, but not in that of question 9.

**14 Recalculate the Top 10 most frequently rated movies. But use only ratings of at least 4 stars. Provide the movie ID, number of users who rated the movie, and title for each.**

| movie ID | freq | title   | genre                               |
|----------|------|---|-------------------------------------|
| 2858     | 2853 | American Beauty (1999)                            | Comedy Drama                        |
| 260      | 2622 | Star Wars: Episode IV - A New Hope (1977)         | Action Adventure Fantasy Sci-Fi     |
| 1196     | 2510 | Star Wars: Episode V - The Empire Strikes Back... | Action Adventure Drama Sci-Fi War   |
| 1198     | 2260 | Raiders of the Lost Ark (1981)                    | Action Adventure                    |
| 2028     | 2260 | Saving Private Ryan (1998)                        | Action Drama War                    |
| 593      | 2252 | Silence of the Lambs, The (1991)                  | Drama Thriller                      |
| 2571     | 2171 | Matrix, The (1999)                                | Action Sci-Fi Thriller              |
| 2762     | 2163 | Sixth Sense, The (1999)                           | Thriller                            |
| 1210     | 2127 | Star Wars: Episode VI - Return of the Jedi (1983) | Action Adventure Romance Sci-Fi War |
| 608      | 2074 | Fargo (1996)                                      | Crime Drama Thriller                |

**15 Compare the resulting lists of question 9 and 14. What do you witness and how can you explain this?**

Most of the popular movies from question 9 are also listed in table of question 14. This is because people are most motivated to rate items when they either loved or hate it. The phenom causes ratings of 5 or 1 and no mediocre ratings. This is a drawback of aggregative rating.