Collaborative Filtering

Generating film ratings

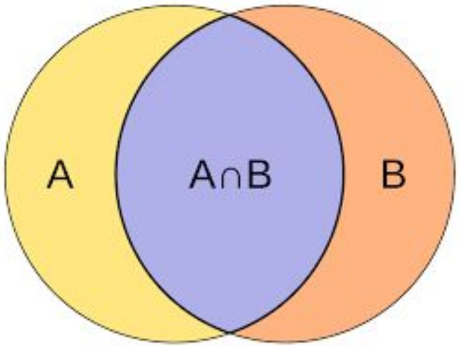
M2 BDMA
Decision Modelling
Fall 2023
Jose Antonio Lorencio Abril

The 5th metric: Jaccard Similarity

Jaccard Similarity

Given two sets, A and B, their Jaccard index is

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$



The 5th metric: Jaccard Similarity

Jaccard Similarity

Given two sets, A and B, their Jaccard index is

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Problem:

- Jaccard does not account for the ratings given by users
- It only accounts for appearance or not

The 5th metric: Tanimoto Similarity

Tanimoto Similarity

Given two vectors, A and B, their Tanimoto index is

$$T(A,B) = \frac{\sum a_i \cdot b_i}{\sum a_i^2 + \sum b_i^2 - \sum a_i \cdot b_i}$$

Tanimoto solves the Jaccard problem!

Generating Ratings

Generating random ratings

1. For i = 1:N_users
   a. Generate user i
   b. For j = 1:N_films
      i. 40% Chance: Not rated
      ii. 60% Chance: Random rating between 1 and 5

- The 40/60 chances increase the probability of having between 30-50% of blanks.

Generating Ratings

Generating valid random ratings

1. Repeat:
   a. Generate a random rating R

   Until **0.3 < blank_percentage(R) < 0.5** AND **existsValidUser(R)**

- blank_percentage(R) returns the percentage of blanks in R
- existsValidUser(R) returns whether there is a user with at least 50% blanks or not

All methods give the same recommendation

Generating a rating giving the same recommendation with all methods to a user

1. Repeat:
    a. Generate valid ratings R
    b. Get valid user from R, U
    c. Compute recommendations for U

    Until all recommendations coincide

All methods give the same recommendation

```
Generating table of critiques...
The table of critiques is:
          Person 1  Person 2  Person 3  Person 4  Person 5  Person 6  Person 7  Person 8  Person 9  Person 10
Movie 1        5.0       NaN       5.0       NaN       NaN       1.0       NaN       NaN       5.0        NaN
Movie 2        1.0       4.0       NaN       NaN       3.0       NaN       NaN       NaN       1.0        NaN
Movie 3        5.0       NaN       2.0       NaN       3.0       3.0       NaN       3.0       NaN        NaN
Movie 6        4.0       3.0       4.0       5.0       NaN       NaN       NaN       5.0       NaN        1.0
Movie 7        1.0       NaN       1.0       1.0       1.0       2.0       NaN       NaN       3.0        NaN
Movie 9        3.0       3.0       NaN       NaN       2.0       NaN       5.0       5.0       5.0        1.0
Movie 14       1.0       NaN       NaN       3.0       NaN       5.0       3.0       4.0       2.0        5.0
Movie 8        NaN       1.0       NaN       3.0       2.0       NaN       2.0       NaN       2.0        1.0
Movie 10       NaN       3.0       NaN       NaN       NaN       NaN       2.0       NaN       NaN        5.0
Movie 12       NaN       2.0       3.0       3.0       NaN       NaN       5.0       4.0       NaN        NaN
Movie 15       NaN       2.0       1.0       1.0       NaN       5.0       3.0       3.0       NaN        2.0
Movie 4        NaN       NaN       2.0       4.0       4.0       3.0       3.0       2.0       1.0        NaN
Movie 5        NaN       NaN       1.0       5.0       2.0       1.0       5.0       NaN       1.0        2.0
Movie 11       NaN       NaN       3.0       NaN       2.0       5.0       1.0       2.0       1.0        2.0
Movie 13       NaN       NaN       2.0       2.0       1.0       4.0       NaN       1.0       NaN        NaN
There are 64 NaNs in the table, that is 42.666666666666667% of the table.
The table of recommendations is:
              Best Best with exp Best with Pearson Best with cosine Best with Tanimoto
Person 1  Movie 12      Movie 12          Movie 12         Movie 12           Movie 12
```

Each method gives a different recommendation

Generating a rating giving different recommendations for each methods to a user

1. Repeat:
    a. Generate valid ratings R
    b. Get valid user from R, U
    c. Compute recommendations for U

    Until all recommendations differ

Each method gives a different recommendation

```
Generating table of critiques...
The table of critiques is:
          Person 1  Person 2  Person 3  Person 4  Person 5  Person 6  Person 7  Person 8  Person 9  Person 10
Movie 1        2.0       3.0       4.0       2.0       4.0       NaN       NaN       NaN       NaN        4.0
Movie 3        5.0       2.0       2.0       5.0       NaN       3.0       NaN       NaN       NaN        NaN
Movie 5        3.0       NaN       2.0       NaN       NaN       NaN       1.0       NaN       4.0        NaN
Movie 6        5.0       3.0       3.0       NaN       NaN       5.0       2.0       4.0       2.0        NaN
Movie 9        4.0       2.0       NaN       NaN       2.0       1.0       NaN       1.0       NaN        3.0
Movie 10       1.0       3.0       1.0       4.0       NaN       4.0       3.0       NaN       NaN        2.0
Movie 12       3.0       4.0       4.0       NaN       NaN       NaN       2.0       3.0       NaN        NaN
Movie 14       2.0       2.0       NaN       5.0       4.0       NaN       NaN       NaN       3.0        5.0
Movie 4        NaN       2.0       3.0       4.0       NaN       NaN       NaN       1.0       2.0        NaN
Movie 8        NaN       1.0       5.0       NaN       NaN       NaN       NaN       NaN       4.0        NaN
Movie 11       NaN       4.0       NaN       4.0       5.0       1.0       2.0       2.0       NaN        NaN
Movie 13       NaN       2.0       NaN       5.0       NaN       NaN       3.0       NaN       NaN        1.0
Movie 15       NaN       2.0       NaN       1.0       NaN       NaN       5.0       3.0       3.0        NaN
Movie 2        NaN       NaN       1.0       NaN       4.0       NaN       NaN       4.0       NaN        4.0
Movie 7        NaN       NaN       1.0       2.0       NaN       1.0       4.0       2.0       5.0        NaN
There are 71 NaNs in the table, that is 47.333333333333336% of the table.
The table of recommendations is:
             Best Best with exp Best with Pearson Best with cosine Best with Tanimoto
Person 5  Movie 8       Movie 7          Movie 12          Movie 6            Movie 3
```