

UNIVERSITEIT GENT

RECOMMENDER SYSTEMS

## Assignment 2 - Personalized Rec

*Tibo Vanheule*

*2022-2023*

## 1 UUCF

### 1.1 What is the Pearson correlation (without significance weighting) between the user with id 1 and the user with id 4?

The Pearson correlation is 0.042136808375910856.

### 1.2 What is the Pearson correlation (with significance weighting) between the user with id 1 and the user with id 4?

The Pearson correlation is 0.021068404187955428.

### 1.3 Compare the answers of question 1 and 2. Explain the difference (if there is a difference).

The second correlation is smaller due to the significance weighting.

### 1.4 Calculate the ratio of the answer of question 1 and the answer of question 2. Explain this value.

The ratio is 2, this is 10 divided by 5. The formula of significance weighting is  $\min(10, \text{ratings})/10$  and 5 ratings.  $5/10=1/2$ .

### 1.5 Consider the user with id = 1 now as target user. Consider the item with id =10 as target item. How many neighbors (with a strict positive similarity) can you find who rated the target item?

There are 30 neighbours.

### 1.6 What are the Top-k neighbors for the user with id = 1 and the item with id =10 as target item? Specify the ids of the neighbors and the similarity with the target user. Order this list in descending order according to the similarity value.

[(390, 0.37976211745379407).  
(312, 0.34778700225454795).  
(353, 0.2865915778360536).  
(428, 0.2729107049454655).  
(177, 0.2630318859067766).  
(561, 0.25583582053934906).  
(405, 0.2023994590198254).  
(458, 0.19999815803167262).  
(236, 0.19012109129819604).  
(247, 0.18298057613606222).  
(430, 0.17698364447639653).  
(243, 0.1627484990010214).  
(285, 0.14863349494711478).  
(574, 0.14297980920771322).

(150, 0.14281903684635436).  
 (608, 0.14008707563881007).  
 (295, 0.13295134132714795).  
 (641, 0.12493900951088494).  
 (592, 0.12403408673701902).  
 (384, 0.12083680180003592)]

**1.7 Consider the user with id = 1 now as target user. Consider the item with id = 10 as target item. Use the ratings of the Top-k neighbors of the user. What is the weighted average of the deviation from their mean rating?**

The weighted average of the deviation from their mean rating is -0.30277403965828986.

**1.8 What is the rating prediction for the user with id = 1 and the item with id = 10?**

The rating prediction is 2.24722596034171.

**1.9 What is the title of the item with id = 10?**

movieID	title	genre
9	10 GoldenEye (1995)	Action Adventure Thriller

**1.10 Consider the user with id = 1 now as target user. Consider the item with id = 260 as target item. How many neighbors (with a strict positive similarity) can you find who rated the target item?**

I found 63 neighbours with a strict positive similarity.

**1.11 What are the Top-k neighbors for the user with id = 1 and the item with id = 260 as target item? Specify the ids of the neighbors and the similarity with the target user. Order this list in descending order according to the similarity value.**

[(580, 0.662547274200393).  
 (463, 0.388763881018101).  
 (390, 0.37976211745379407).  
 (585, 0.35062100531729185).  
 (312, 0.34778700225454795).  
 (353, 0.2865915778360536).  
 (394, 0.2816898480519987).  
 (466, 0.28001045251295337).  
 (510, 0.2767346746437224).  
 (428, 0.2729107049454655).  
 (505, 0.25694970235409637).  
 (561, 0.25583582053934906).  
 (533, 0.2485524986721755).

(242, 0.24017846543583932).  
(41, 0.2179899058684344).  
(22, 0.21035896427251233).  
(439, 0.20804249079711562).  
(405, 0.2023994590198254).  
(452, 0.19218775874226243).  
(236, 0.19012109129819604)]

**1.12 Consider the user with id = 1 now as target user. Consider the item with id = 260 as target item. Use the ratings of the Top-k neighbors of the user. What is the weighted average of the deviation from their mean rating?**

The weighted average of the deviation from their mean rating is 0.5054658898192962

**1.13 What is the rating prediction for the user with id = 1 and the item with id = 260?**

the rating prediction is 3.055465889819296.

**1.14 What is the title of the item with id = 260?**

movieID	title	genre
232	260 Star Wars: Episode IV - A New Hope (1977)	Action Adventure Sci-Fi

**1.15 Compare the lists of question 6 and 11. What can you say about this? Explain the similarities / differences.**

movie 10 has 122 user rating and movie 260 has 291 ratings. 2 of the 3 genres are the same (action and adventure). But they have a different third genre, sci-fi vs thriller.

**1.16 Calculate the Top-N recommendations for the user with id = 1. Specify item id, title, and rating prediction in a list. Order the list according to the rating prediction in descending order.**

movieID	rating prediction	title	genre
0	3216	3216 Vampyros Lesbos (Vampirás, Las) (1971)	Fantasy Horror Thriller
1	40412	40412 Dead Man's Shoes (2004)	Crime Thriller
2	92494	92494 Dylan Moran: Monster (2004)	Comedy Documentary
3	3320	3320 Mifune's Last Song (Mifunes sidste sang) (1999)	Comedy Drama Romance
4	4302	4302 King Is Alive, The (2000)	Drama
5	4731	4731 Innocence (2000)	Drama
6	5071	5071 Maelström (2000)	Drama Romance
7	86781	86781 Incendies (2010)	Drama Mystery War
8	97957	97957 Excision (2012)	Crime Drama Horror Thriller
9	3414	3414 Love Is a Many-Splendored Thing (1955)	Drama Romance War

- 1.17 Calculate the Top-N recommendations for the user with id = 522. Specify item id, title, and rating prediction in a list. Order the list according to the rating prediction in descending order.**

	movieID	rating prediction	title	genre
0	565	565	Cronos (1993)	Drama Horror
1	1450	1450	Prisoner of the Mountains (Kavkazsky plennik) ...	War
2	1563	1563	Dream With the Fishes (1997)	Drama
3	1819	1819	Storefront Hitchcock (1997)	Documentary Musical
4	4076	4076	Two Ninas (1999)	Comedy Romance
5	4591	4591	Erik the Viking (1989)	Adventure Comedy Fantasy
6	4796	4796	Grass Is Greener, The (1960)	Comedy Romance
7	4930	4930	Funeral in Berlin (1966)	Action Drama Thriller
8	5427	5427	Caveman (1981)	Comedy
9	3216	3216	Vampyros Lesbos (Vampirás, Las) (1971)	Fantasy Horror Thriller

## 2 IICF

- 2.1 For this implementation of IICF, we do not use an additional factor for significance weighting, as we did for UUCF. Explain why such an additional factor is not useful here.**

This isn't useful here since denominator uses all the ratings and not just the ratings from the common users.

- 2.2 Create your item-item model. How many strict positive similarities are stored in your model?**

4393578 similarities are saved.

- 2.3 What is the cosine similarity between the items with id = 594 and id = 596? Take a look at the movie titles. Did you expect such a value for these 2 movies? Why?**

Movies have quite high similarity of 0.3104323318682939, but since the genre and time area are the same this is to be expected.

- 2.4 Consider now the user with id=522 as the target user. The goal is to calculate a rating prediction for the target item with id=25. Therefore, we have to identify items that are rated by the target user, and are similar to the target item. How many of these items with a similarity >0 can be found?**

160 items with a positive similarity can be found.

- 2.5** Consider now the user with id=522 as the target user and the item with id=25 as the target item. Provide a list of the top-k most similar items (rated by the target user), together with the similarity score. The list has to be ordered according to the score in descending order. Provide ids, titles, and similarity scores.

	cosine	movieID	title	genre
0	0.146787	4239.0	Blow (2001)	Crime Drama
1	0.138505	44191.0	V for Vendetta (2006)	Action Sci-Fi Thriller IMAX
2	0.138027	778.0	Trainspotting (1996)	Comedy Crime Drama
3	0.130573	1092.0	Basic Instinct (1992)	Crime Mystery Thriller
4	0.124621	485.0	Last Action Hero (1993)	Action Adventure Comedy Fantasy
5	0.121433	81845.0	King's Speech, The (2010)	Drama
6	0.114517	165.0	Die Hard: With a Vengeance (1995)	Action Crime Thriller
7	0.104864	122.0	Boomerang (1992)	Comedy Romance
8	0.098115	4019.0	Finding Forrester (2000)	Drama
9	0.097993	68157.0	Inglourious Basterds (2009)	Action Drama War

- 2.6** Compare the values of the similarities in the list of question 22 (cosine similarity) with the values of the similarities in question 6 and 11 (Pearson correlation). What can you conclude here?

The maximum score achieved with cosine (0.146787) is much lower than the maximum of Pearson correlation (0.66).

- 2.7** Calculate the Top-N recommendations for the user with id = 522. Use only positive similarities. Specify item id, title, and rating prediction in a list. Order the list according to the rating prediction in descending order.

	movieID	rating prediction	title	genre
0	1232.0	3.912456	Stalker (1979)	Drama Mystery Sci-Fi
1	5690.0	3.894039	Grave of the Fireflies (Hotaru no haka) (1988)	Animation Drama War
2	58998.0	3.879760	Forgetting Sarah Marshall (2008)	Comedy Romance
3	6794.0	3.873757	Beethoven's 2nd (1993)	Children Comedy
4	69275.0	3.848123	Dead Snow (Død snø) (2009)	Action Adventure Comedy Horror
5	72407.0	3.845033	Twilight Saga: New Moon, The (2009)	Drama Fantasy Horror Romance Thriller
6	62336.0	3.830744	FLCL (2000)	Animation Comedy Fantasy Sci-Fi
7	5903.0	3.822589	Equilibrium (2002)	Action Sci-Fi Thriller
8	1088.0	3.811169	Dirty Dancing (1987)	Drama Musical Romance
9	37475.0	3.809554	Unfinished Life, An (2005)	Drama

### 3 Basket

- 3.1 Calculate the Top-N recommendations for a shopping basket with one item. The item in the basket is the item with id=1. Only positive similarities have to be used. Order the list according to the output of the recommender in descending order. Are the results what you expect? Why?**

	movieID	rating	prediction	title	genre
0	3114.0	0.387519		Toy Story 2 (1999)	Adventure Animation Children Comedy Fantasy
1	2355.0	0.291851		Bug's Life, A (1998)	Adventure Animation Children Comedy
2	364.0	0.267971		Lion King, The (1994)	Adventure Animation Children Drama Musical IMAX
3	588.0	0.242452		Aladdin (1992)	Adventure Animation Children Comedy Musical
4	4886.0	0.236092		Monsters, Inc. (2001)	Adventure Animation Children Comedy Fantasy
5	4306.0	0.233792		Shrek (2001)	Adventure Animation Children Comedy Fantasy Ro...
6	58559.0	0.222851		Dark Knight, The (2008)	Action Crime Drama IMAX
7	5349.0	0.222540		Spider-Man (2002)	Action Adventure Sci-Fi Thriller
8	1092.0	0.221361		Basic Instinct (1992)	Crime Mystery Thriller
9	8961.0	0.220612		Incredibles, The (2004)	Action Adventure Animation Children Comedy

	movieID	title	genre
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy

Yes, the movies are in line with Toy story (id=1). They have much genres in common and most recommendations are also for children.

- 3.2 Calculate the Top-N recommendations for a shopping basket with the following items: the items with id=1, id=48, id=239. Only positive similarities have to be used. Order the list according to the output of the recommender in descending order. Are the results what you expect? Why?**

	movieID	rating	prediction	title	genre
0	2355.0	0.633219		Bug's Life, A (1998)	Adventure Animation Children Comedy
1	3810.0	0.590816		White Sands (1992)	Drama Thriller
2	364.0	0.585337		Lion King, The (1994)	Adventure Animation Children Drama Musical IMAX
3	3807.0	0.585142		Sinbad and the Eye of the Tiger (1977)	Adventure Fantasy
4	2089.0	0.575403		Rescuers Down Under, The (1990)	Adventure Animation Children
5	54276.0	0.532559		No Reservations (2007)	Comedy Drama Romance
6	2772.0	0.526674		Detroit Rock City (1999)	Comedy
7	3114.0	0.499952		Toy Story 2 (1999)	Adventure Animation Children Comedy Fantasy
8	2531.0	0.493176		Battle for the Planet of the Apes (1973)	Action Sci-Fi
9	2045.0	0.492943		Far Off Place, A (1993)	Adventure Children Drama Romance

	movieID	title	genre
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy
46	48	Pocahontas (1995)	Animation Children Drama Musical Romance
211	239	Goofy Movie, A (1995)	Animation Children Comedy Romance

The movies are more mixed, the genre fantasy is less frequent. And common genres are still very good represented.

**3.3 Calculate the Top-N recommendations for a shopping basket with the following items: the items with id=1, id=48, id=239. Positive as well as negative similarities have to be used. Order the list according to the output of the recommender in descending order. Are the results what you expect? Why?**

	movieID	rating prediction	title	genre
0	2355.0	0.633219	Bug's Life, A (1998)	Adventure Animation Children Comedy
1	364.0	0.585337	Lion King, The (1994)	Adventure Animation Children Drama Musical IMAX
2	2089.0	0.575403	Rescuers Down Under, The (1990)	Adventure Animation Children
3	54276.0	0.518988	No Reservations (2007)	Comedy Drama Romance
4	3114.0	0.499952	Toy Story 2 (1999)	Adventure Animation Children Comedy Fantasy
5	2772.0	0.414240	Detroit Rock City (1999)	Comedy
6	3807.0	0.412500	Sinbad and the Eye of the Tiger (1977)	Adventure Fantasy
7	3810.0	0.406068	White Sands (1992)	Drama Thriller
8	2045.0	0.360498	Far Off Place, A (1993)	Adventure Children Drama Romance
9	2531.0	0.357576	Battle for the Planet of the Apes (1973)	Action Sci-Fi

**3.4 Compare the results of question 26 and 27. What can you say about this? Which list is the best?**

Most movies of question 3.3 are in question 3.4, but some genres are less dominant. I prefer the method with both positive and negative similarities.

## 4 Hybrid

**4.1 Calculate the Top-N recommendations for the user with id 522 using the Hybrid recommender. Only positive similarities have to be used. Specify item id, title, and rating prediction in a list. Order the list according to the rating prediction in descending order.**

	score	movieID	title	genre
0	4.785880	565	Cronos (1993)	Drama Horror
1	4.489903	40412	Dead Man's Shoes (2004)	Crime Thriller
2	4.120966	2304	Love Is the Devil (1998)	Drama
3	4.094669	4518	The Lair of the White Worm (1988)	Comedy Horror
4	4.085897	85438	Jane Eyre (2011)	Drama Romance
5	4.072171	85179	Summer Wars (Samâ wôzu) (2009)	Adventure Animation Comedy Sci-Fi
6	4.071368	6063	May (2002)	Drama Horror
7	4.059823	26587	Decalogue, The (Dekalog) (1989)	Crime Drama Romance
8	4.049276	1564	For Roseanna (Roseanna's Grave) (1997)	Comedy Drama Romance
9	4.040194	1860	Character (Karakter) (1997)	Drama

**4.2 Compare the hybrid recommendations of question 29 with the answers of 17 and 24. What can you say about this.**

all tables contains different movies. hybrid is a nice compromise.