

Assignment 4: Social Recommendation

Date of issue: 28.03.2019

Natalia Ostapuk: natalia.ostapuk@unifr.ch

Task 1. Table 1 contains users' ratings for 4 movies. Predict the missing rating using the item based collaborative filtering. Use $k=2$ as the neighborhood size and the cosine similarity measure.

	Forrest Gump	Intouchables	Fight Club	Lion King
Michael	3	4	5	3
Paul	5	5	3	5
Ann	5	?	2	5
Julie	5	4	3	4
Pierre	4	3	1	3
Sophie	4	5	4	5

Table 1

Task 2. Table 2 contains users' ratings for 5 movies, some ratings are missed. Implement the user based collaborative filtering to predict the missing values.

	Forrest Gump	Intouchables	Fight Club	Lion King	Pulp Fiction
Michael	?	4	5	3	5
Paul	5	5	3	5	4
Ann	5	?	2	5	3
Julie	5	4	3	4	?
Pierre	4	3	?	3	5
Sophie	4	5	4	5	5

Table 2

Hints:

- Store the data into a dictionary, in which the keys represent the users and the values are dictionaries containing movies and ratings.
- In order to predict a missing rate, first compute the average ratings for the users, then use the Pearson coefficient to measure the similarity between two users and finally compute the missing ratings for each user.
- Use $k=2$ as the neighborhood size.
- If the predicted score is smaller than 0, then set it to 0, and if the predicted score is bigger than 5, then set it to 5.

Task 3. Implement the item based collaborative filtering to predict the missing values in the dataset of the Task 2.

Hints:

- Store the data into a dictionary. Note that it would be more convenient to store the movies as keys, and users with respective ratings as values.
- In order to predict a missing rate, first compute the average ratings for the movies, then use the Pearson coefficient to measure the similarity between two movies and finally compute the missing ratings for each movie.
- Use $k=2$ as the neighborhood size.
- If the predicted score is smaller than 0, then set it to 0, and if the predicted score is bigger than 5, then set it to 5.

Assignment delivery notes:

- This assignment should be uploaded to ILIAS by 03.04.2019
- Task 1 should be uploaded in *pdf* format.
- Tasks 2 and 3 should be uploaded in *py* format.