

Assignment 2: Group Recommendations

Group members: Tuomas Porkamaa, Mueed Irfan

We Implemented solutions for all subtasks.

How to run the code:

The submission consist a single python file:

- *Assignment2.py*

Program can be executed in terminal by a command `<python Assignment2.py>`. However, the suggested environment for running the program is the editor environment of Spyder IDE. There you can run parts of the program (implemented as cells) separately, and the program outputs make more sense and you can also modify user-specified hyperparameters more easily.

NOTE BEFORE RUNNING THE CODE:

Modify the root variable to point to the directory (explicitly) where the program data is located (ratings.csv and movies.csv). Current setting is

```
root = '/home/tuomas/Python/DATA.ML.360/Assignment2/ml-latest-small/'
```

This variable can be found in following line:

- *Assignment2.py*, line 9.

Also, if you want to change the current user group under examination (current group initialized to [1,2,3]), change the following line:

- *Assignment2.py*, line 127.

In part B, you can control the 'disagreement level' by modifying the second parameter of the *calc_weights(a,b)* -function. Current setup is *calc_weights(a,2)*. You can modify this in the following line:

- *Assignment2.py*, line 220.

Assumptions

Same assumptions hold that were noted in the manual of assingment 1 (*Assignment1_manual.pdf*). Also, in the group recommendation part, we considered only the common unrated movies between group members.

Summary of implementation in part B

In part B, the given task was to define a way for counting the disagreements between the users in a group, and propose a method that takes disagreements into account when computing suggestions for the group. Our implementation's mathematical background is following:

Sample variance of the predicted ratings is intuitive method to measure the disagreements between group members. That is the main building block of our method

$$Var(\mathbf{r}_i) = \frac{1}{|G|} \sum_{u \in G} (r_{ui} - mean(\mathbf{r}_i))^2 \quad (\text{Equation 1})$$

In Eq 1.

- \mathbf{r}_i = ratings of common unrated item i
- G = group of users
- r_{ui} = user u 's predicted rating for item I

Next step is to calculate the weightings for common rating metrics (eg. average in group). This is done as follows:

$$w_i = \frac{(Var(\mathbf{r}_i) + \epsilon)^{-1}}{\sum_k (Var(\mathbf{r}_k) + \epsilon)^{-1}} \quad (\text{Equation 2})$$

The final ratings are calculated by $w_i * mean(\mathbf{r}_i)$ and the item i 's with largest values are suggested for the group. Weighting can be also applied to other rating metrics as well (eg. the ones calculated by Least misery method). In addition, be aware that the new predicted scores doesn't make sense anymore. They are only used for comparison..

In Eq 2.

- Large variance (large disagreement) should be punished somehow. This is done by weighting the samples by inverse of the variance.
- Denominator guarantees that w_i 's sum to one. This is just for personal preferences.
- Parameter ϵ controls the role of the variance in weighting. By setting $\epsilon = 0$ the variance (disagreement level) has a large effect in the final comparison process. By setting ϵ to a larger value (eg. 10) the variance has a minimal effect in the final comparison process and the actual rating scores are in the bigger role.

Example outputs from the program.

Group recommendations using group average.

List of 20 most relevant movies for user group [1 2 3]:

(Id : 5490), The Big Bus (1976)

(Id : 132333), Seve (2014)

(Id : 25947), Unfaithfully Yours (1948)

(Id : 3379), On the Beach (1959)

(Id : 8477), Jetée, La (1962)

(Id : 67618), Strictly Sexual (2008)

(Id : 6818), Come and See (Idi i smotri) (1985)

(Id : 3086), Babes in Toyland (1934)

(Id : 162344), Tom Segura: Mostly Stories (2016)

(Id : 128914), Tom Segura: Completely Normal (2014)

(Id : 152105), Dad's Army (1971)

(Id : 7926), High and Low (Tengoku to jigoku) (1963)

(Id : 50610), Beer League (2006)

(Id : 25906), Mr. Skeffington (1944)

(Id : 77846), 12 Angry Men (1997)

(Id : 93008), Very Potter Sequel, A (2010)

(Id : 107780), Cats (1998)

(Id : 26326), Holy Mountain, The (Montaña sagrada, La) (1973)

(Id : 5328), Rain (2001)

(Id : 3951), Two Family House (2000)

Group recommendations using least misery method.

List of 20 most relevant movies for user group [1 2 3]:

(Id : 40491), Match Factory Girl, The (Tulitikkutehtaan tyttö) (1990)

(Id : 3567), Bossa Nova (2000)

(Id : 156605), Paterson

(Id : 136850), Villain (1971)

(Id : 3940), Slumber Party Massacre III (1990)

(Id : 3941), Sorority House Massacre (1986)

(Id : 3942), Sorority House Massacre II (1990)

(Id : 3939), Slumber Party Massacre II (1987)

(Id : 3851), I'm the One That I Want (2000)

(Id : 1739), 3 Ninjas: High Noon On Mega Mountain (1998)

(Id : 57502), Cat Soup (Nekojiru-so) (2001)

(Id : 1140), Entertaining Angels: The Dorothy Day Story (1996)

(Id : 7096), Rivers and Tides (2001)

(Id : 148), Awfully Big Adventure, An (1995)

(Id : 496), What Happened Was... (1994)

(Id : 141718), Deathgasm (2015)

(Id : 3795), Five Senses, The (1999)

(Id : 81535), Saw VII 3D - The Final Chapter (2010)

(Id : 7184), This Property is Condemned (1966)

(Id : 40), Cry, the Beloved Country (1995)

Group recommendations using new method.

List of 20 most relevant movies for user group [1 2 3]:

(Id : 5490), The Big Bus (1976)

(Id : 132333), Seve (2014)

(Id : 25947), Unfaithfully Yours (1948)

(Id : 5915), Victory (a.k.a. Escape to Victory) (1981)

(Id : 89759), Separation, A (Jodaeiye Nader az Simin) (2011)

(Id : 183897), Isle of Dogs (2018)

(Id : 3379), On the Beach (1959)

(Id : 85), Angels and Insects (1995)

(Id : 67618), Strictly Sexual (2008)

(Id : 1437), Cement Garden, The (1993)

(Id : 7841), Children of Dune (2003)

(Id : 5867), Thief (1981)

(Id : 199), Umbrellas of Cherbourg, The (Parapluies de Cherbourg, Les) (1964)

(Id : 2295), Impostors, The (1998)

(Id : 3086), Babes in Toyland (1934)

(Id : 73344), Prophet, A (Un Prophète) (2009)

(Id : 69524), Raiders of the Lost Ark: The Adaptation (1989)

(Id : 86377), Louis C.K.: Shameless (2007)

(Id : 162344), Tom Segura: Mostly Stories (2016)

(Id : 128914), Tom Segura: Completely Normal (2014)