Assignment 8

CS595

Introduction to Web Science

Old Dominion University

Computer Science

Due: 11:59 pm Nov 14

Lulwah Alkwai

Question 1

The goal of this project is to use the basic recommendation principles we have learned for user-collected data. You will modify the code given to you which performs movie recommendations from the MovieLense data sets.

The MovieLense data sets were collected by the GroupLens Research Project at the University of Minnesota during the seven-month period from September 19th, 1997 through April 22nd, 1998. It is available for download from `http://www.grouplens.org/node/73`

There are three files which we will use:

1. u.data: 100,000 ratings by 943 users on 1,682 movies. Each user has rated at least 20 movies. Users and items are numbered consecutively from 1. The data is randomly ordered. This is a tab separated list of

user id — item id — rating — timestamp

The time stamps are unix seconds since 1/1/1970 UTC.

Example:

196 242 3 881250949
186 302 3 891717742
22 377 1 878887116
244 51 2 880606923
166 346 1 886397596
298 474 4 884182806
115 265 2 881171488

2. u.item: Information about the 1,682 movies. This is a tab separated list of

```
movie id | movie title | release date | video release date |
IMDb URL | unknown | Action | Adventure | Animation |Children's |
Comedy | Crime | Documentary | Drama | Fantasy | Film-Noir |
Horror | Musical | Mystery | Romance | Sci-Fi | Thriller | War |
Western |
```

The last 19 fields are the genres, a 1 indicates the movie is of that genre,

a 0 indicates it is not; movies can be in several genres at once. The movie ids are the ones used in the u.data data set.

Example:

```
161|Top Gun (1986)|01-Jan-1986||http://us.imdb.com/M/title-exact?
Top%20Gun%20(1986)
|0|1|0|0|0|0|0|0|0|0|0|0|0|0|0|1|0|0|0|0
162|On Golden Pond (1981)|01-Jan-1981||
http://us.imdb.com/M/title-exact?On%20Golden%20Pond%20(1981)
|0|0|0|0|0|0|0|0|0|1|0|0|0|0|0|0|0|0|0|0
163|Return of the Pink Panther, The (1974)|01-Jan-1974||
http://us.imdb.com/M/title-exact?Return%20of
%20the%20Pink%20Panther,%20The%20(1974)
|0|0|0|0|0|1|0|0|0|0|0|0|0|0|0| 0|0|0|0|0
```

3. u.user: Demographic information about the users. This is a tab separated list of:

```
user id | age | gender | occupation | zip code
```

The user ids are the ones used in the u.data data set.

Example:

```
1|24|M|technician|85711\\
2|53|F|other|94043\\
3|23|M|writer|32067\\
4|24|M|technician|43537\\
5|33|F|other|15213\\
```

The code for reading from the u.data and u.item files and creating recommendations is described in the book Programming Collective Intelligence (check email for more details). You are to modify recommendations.py to answer the following questions. Each question your program answers correctly will award you 10 points. You must have the question answered completely correct; partial credit will only be awarded if your answer is very close to the correct one.

1. What 5 movies have the highest average ratings? Show the movies and their ratings sorted by their average ratings.

2. What 5 movies received the most ratings? Show the movies and the number of ratings sorted by number of ratings.

3. What 5 movies were rated the highest on average by women? Show the movies and their ratings sorted by ratings.

4. What 5 movies were rated the highest on average by men? Show the movies and their ratings sorted by ratings.

5. What movie received ratings most like Top Gun? Which movie received ratings that were least like Top Gun (negative correlation)?

6. Which 5 raters rated the most films? Show the raters' IDs and the number of films each rated.

7. Which 5 raters most agreed with each other? Show the raters' IDs and Pearson's r, sorted by r.

8. Which 5 raters most disagreed with each other (negative correlation)? Show the raters' IDs and Pearson's r, sorted by r.

9. What movie was rated highest on average by men over 40? By men under 40?

10. What movie was rated highest on average by women over 40? By women under 40?

Your output should clearly indicate the answers from the question you answered. Provide any relevant discussion.

Answers

The code is in the altered version of recommendations.py. The output of the code is listed below.

1. What 5 movies have the highest average ratings? Show the movies and their ratings sorted by their average ratings.

('They Made Me a Criminal (1939)', 5.0)
("Someone Else's America (1995)", 5.0)
('Saint of Fort Washington, The (1993)', 5.0)
('Star Kid (1997)', 5.0)
('Entertaining Angels: The Dorothy Day Story (1996)', 5.0)
('Marlene Dietrich: Shadow and Light (1996) ', 5.0)
('Aiqing wansui (1994)', 5.0)
('Prefontaine (1997)', 5.0)
('Great Day in Harlem, A (1994)', 5.0)
('Santa with Muscles (1996)', 5.0)

2. What 5 movies received the most ratings? Show the movies and the number of ratings sorted by number of ratings.

1– ('Star Wars (1977)', 583)
2– ('Contact (1997)', 509)
3– ('Fargo (1996)', 508)
4– ('Return of the Jedi (1983)', 507)
5– ('Liar Liar (1997)', 485)

3. What 5 movies were rated the highest on average by women? Show the movies and their ratings sorted by ratings.

('Stripes (1981)', 5.0)
("Someone Else's America (1995)", 5.0)
('Faster Pussycat! Kill! Kill! (1965)', 5.0)
('Everest (1998)', 5.0)
('Visitors, The (Visiteurs, Les) (1993)', 5.0)
('Maya Lin: A Strong Clear Vision (1994)', 5.0)
('Mina Tannenbaum (1994)', 5.0)
('Year of the Horse (1997)', 5.0)
('Foreign Correspondent (1940)', 5.0)
('Telling Lies in America (1997)', 5.0)
('Prefontaine (1997)', 5.0)

4. What 5 movies were rated the highest on average by men? Show
    the movies and their ratings sorted by ratings.

('They Made Me a Criminal (1939)', 5.0)
('Letter From Death Row, A (1998)', 5.0)
('Saint of Fort Washington, The (1993)', 5.0)
('Quiet Room, The (1996)', 5.0)
('Star Kid (1997)', 5.0)
('Entertaining Angels: The Dorothy Day Story (1996)', 5.0)
('Marlene Dietrich: Shadow and Light (1996) ', 5.0)
('Little City (1998)', 5.0)
('Aiqing wansui (1994)', 5.0)
('Prefontaine (1997)', 5.0)
('Love Serenade (1996)', 5.0)
('Leading Man, The (1996)', 5.0)
('Great Day in Harlem, A (1994)', 5.0)
('Delta of Venus (1994)', 5.0)
('Santa with Muscles (1996)', 5.0)

5. What movie received ratings most like Top Gun?

Shiloh (1997)

Part2−5. Which movie received ratings that were least like Top
    Gun (negative correlation)?

Babysitter, The (1995)

6. Which 5 raters rated the most films? Show the raters' IDs
    and the number of films each rated.

1− ('405', 736)
2− ('655', 678)
3− ('13', 632)
4− ('450', 538)
5− ('276', 516)

7. Which 5 raters most agreed with each other? Show the raters
    IDs and Pearson's r, sorted by r.

781,477,1.0\
781,754,1.0\
781,939,1.0\
381,309,1.0\
361,570,1.0\
361,571,1.0\

8. Which 5 raters most disagreed with each other (negative
    correlation)? Show the raters' IDs and Pearson's r, sorted by

r.

782,225,−1.0\
783,908,−1.0\
789,909,−1.0\
860,602,−1.0\
579,675,−1.0\

9.   What movie was rated highest on average by men over 40?

('Spice World (1997)', 5.0)
('Indian Summer (1996)', 5.0)
('Two or Three Things I Know About Her (1966)', 5.0)
('Unstrung Heroes (1995)', 5.0)
('Boxing Helena (1993)', 5.0)
('Poison Ivy II (1995)', 5.0)
('They Made Me a Criminal (1939)', 5.0)
('Star Kid (1997)', 5.0)
('Little City (1998)', 5.0)
('Great Day in Harlem, A (1994)', 5.0)
('Leading Man, The (1996)', 5.0)
('Late Bloomers (1996)', 5.0)
('Strawberry and Chocolate (Fresa y chocolate) (1993)', 5.0)
('Little Princess, The (1939)', 5.0)
('Double Happiness (1994)', 5.0)
('Aparajito (1956)', 5.0)
('Solo (1996)', 5.0)
('Ace Ventura: When Nature Calls (1995)', 5.0)
('Rendezvous in Paris (Rendez−vous de Paris, Les) (1995)', 5.0)
('World of Apu, The (Apur Sansar) (1959)', 5.0)
('Prefontaine (1997)', 5.0)
('Grateful Dead (1995)', 5.0)
('Marlene Dietrich: Shadow and Light (1996) ', 5.0)
('Faithful (1996)', 5.0)
('Hearts and Minds (1996)', 5.0)

Part2−9.   By men under 40?

('Letter From Death Row, A (1998)', 5.0)
('Perfect Candidate, A (1996)', 5.0)
('Saint of Fort Washington, The (1993)', 5.0)
('Quiet Room, The (1996)', 5.0)
('Magic Hour, The (1998)', 5.0)
('Star Kid (1997)', 5.0)
('Entertaining Angels: The Dorothy Day Story (1996)', 5.0)
('Maya Lin: A Strong Clear Vision (1994)', 5.0)
('Angel Baby (1995)', 5.0)
('Love in the Afternoon (1957)', 5.0)

('Aiqing wansui (1994)', 5.0)
('Prefontaine (1997)', 5.0)
('Love Serenade (1996)', 5.0)
('Leading Man, The (1996)', 5.0)
('Crossfire (1947)', 5.0)
('Santa with Muscles (1996)', 5.0)
('Delta of Venus (1994)', 5.0)

10. What movie was rated highest on average by women over 40?

('Safe (1995)', 5.0)
('Pocahontas (1995)', 5.0)
('Grand Day Out, A (1992)', 5.0)
('Nightmare Before Christmas, The (1993)', 5.0)
('Gold Diggers: The Secret of Bear Mountain (1995)', 5.0)
("Mary Shelley's Frankenstein (1994)", 5.0)
('Top Hat (1935)', 5.0)
('Mina Tannenbaum (1994)', 5.0)
('Letter From Death Row, A (1998)', 5.0)
('Band Wagon, The (1953)', 5.0)
('Great Dictator, The (1940)', 5.0)
('Angel Baby (1995)', 5.0)
('Bride of Frankenstein (1935)', 5.0)
('Tombstone (1993)', 5.0)
('Quest, The (1996)', 5.0)
('Funny Face (1957)', 5.0)
('Wrong Trousers, The (1993)', 5.0)
('Balto (1995)', 5.0)
('Foreign Correspondent (1940)', 5.0)
('Best Men (1997)', 5.0)
('Swept from the Sea (1997)', 5.0)
('Ma vie en rose (My Life in Pink) (1997)', 5.0)
('Visitors, The (Visiteurs, Les) (1993)', 5.0)
('In the Bleak Midwinter (1995)', 5.0)
('Shall We Dance? (1937)', 5.0)
('Shallow Grave (1994)', 5.0)

Part2−10. By women under 40?

('Stripes (1981)', 5.0)
("Don't Be a Menace to South Central While Drinking Your Juice
    in the Hood (1996)", 5.0)
("Someone Else's America (1995)", 5.0)
('Grace of My Heart (1996)', 5.0)
('Horseman on the Roof, The (Hussard sur le toit, Le) (1995)',
    5.0)
('Faster Pussycat! Kill! Kill! (1965)', 5.0)

```
('Wedding Gift, The (1994)', 5.0)
("Heaven's Prisoners (1996)", 5.0)
('Everest (1998)', 5.0)
('Umbrellas of Cherbourg, The (Parapluies de Cherbourg, Les)
    (1964)', 5.0)
('Nico Icon (1995)', 5.0)
('Maya Lin: A Strong Clear Vision (1994)', 5.0)
('Mina Tannenbaum (1994)', 5.0)
('Year of the Horse (1997)', 5.0)
('Telling Lies in America (1997)', 5.0)
('Prefontaine (1997)', 5.0)
('Backbeat (1993)', 5.0)
```