CS 495 - Introduction to Web Science

Fall 2014

Assignment 8

by

Eric Littley
UIN: 00821698

November 13, 2014

Instructor

Dr. Michael Nelson

Department of Computer Science
Old Dominion University

Contents

1 Introduction

The purpose of this assignment was to utilize some of the techniques taught in class for making good recommendations to the users of software in which a rating scheme is implemented.

2 Design

The data used in this assignment is from the MovieLens data set [**?**]. The "recommendations.py" code used for part of this assignment is from the "Programming Collective Intelligence" [**?**].

There were three files from the MovieLens data set: u.data, u.user, and u.item. These contain the user ratings, user information, and movie information respectively. These files are parsed and the information is stored in a list of objects, the objects stored are movie objects that contain the information for a specific movie along with it's average rating and number of ratings it received. This processes of computing the averages takes a long time (partially due to my poorly realized design) so this process is split from the main program for this assignment. The program responsible for parsing and averaging the movies is called "parsefiles.py" this program persists the list of movie object and user objects using pickle. "application.py" contains the code to answer each of the questions asked in this assignment. "movies.py" contains class definitions and functions used by both "application.py" and "movies.py." Questions 5, 8, and 9 was answered using code from "recommendations.py" in combination with code that I wrote see the Questions section for more detail.

3 Questions

3.1 Question 1

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

For this question, any movie with less than three ratings was ignored because some movies only have one or two ratings so they have extremely high or low ratings depending on what those raters thought of the movie. The following data shows the top rated movies using this approach.

```
1.  4.63  Pather Panchali (1955)
2.  4.50  Maya Lin: A Strong Clear Vision (1994)
3.  4.49  Close Shave, A (1995)
4.  4.47  Wrong Trousers, The (1993)
5.  4.47  Schindler's List (1993)
```

3.2 Question 2

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

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

3.3 Question 3

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

1. 4.63 Schindler's List (1993)
2. 4.63 Close Shave, A (1995)
3. 4.56 Shawshank Redemption, The (1994)
4. 4.53 Wallace & Gromit: The Best of Aardman Animation (1996)
5. 4.53 Shall We Dance? (1996)

3.4 Question 4

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

1. 4.63 Pather Panchali (1955)
2. 4.5 A Chef in Love (1996)
3. 4.47 Wrong Trousers, The (1993)
4. 4.47 Casablanca (1942)
5. 4.46 Close Shave, A (1995)

3.5 Question 5

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

3.6 Question 6

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

1. 737 405
2. 685 655
3. 636 13
4. 540 450
5. 518 276

3.7 Question 7

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

3.8 Question 8

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

3.9 Question 9

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

3.10 Question 10

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

References