Week 9 Report

Recommendation System

37237256 薛丁銘

I did not do much in this week's work. Just follow the contents taught in the class.
When I selected the movie, I chose to see the most popular and highly rated ones ahead. The code and the result are as shown in Figure 1.

```
print(ratings_average.sort_values(by = ['ratings_count', 'rating'], ascending = False))
```

```
                                       rating  ratings_count
title
Forrest Gump (1994)                  4.164134            329
Shawshank Redemption, The (1994)     4.429022            317
Pulp Fiction (1994)                  4.197068            307
Silence of the Lambs, The (1991)     4.161290            279
Matrix, The (1999)                   4.192446            278
...                                       ...            ...
Wasp Woman, The (1959)               0.500000              1
While the City Sleeps (1956)         0.500000              1
Wizards of the Lost Kingdom II (1989) 0.500000             1
Yongary: Monster from the Deep (1967) 0.500000             1
Zombie Strippers! (2008)             0.500000              1
```

Figure 1. The most popular and highly rated movies

Great! The Matrix! That's it! The recommendation code is as shown in Figure 2 and the result is as shown in Figure 3.

```python
import matplotlib.pyplot as pyplot
import numpy as numpy
import pandas as pd

rattings = pd.read_csv('D:/code/school/Introduction to Machine Learning/Week09_exercise/ml-latest-small/ratings.csv')
movies = pd.read_csv('D:/code/school/Introduction to Machine Learning/Week09_exercise/ml-latest-small/movies.csv')

data = pd.merge(rattings, movies, on = 'movieId')

ratings_average = pd.DataFrame(data.groupby('title')['rating'].mean())
ratings_average['ratings_count'] = pd.DataFrame(data.groupby('title')['rating'].count())

ratings_matrix = data.pivot_table(index='userId', columns='title', values='rating')

# print(ratings_average.sort_values('ratings_count', ascending=False).head(10))
# print(ratings_matrix.head(10))

## find user rating for a Movie
# favorite_movie_ratings = ratings_matrix['Aladdin (1992)']
favorite_movie_ratings = ratings_matrix['Matrix, The (1999)']

#print(favorite_movie_ratings.head(10))

##Finding similar movies
similar_movies = ratings_matrix.corrwith(favorite_movie_ratings)
# print(similar_movies.head(10))

##Remove empty values
correlation = pd.DataFrame(similar_movies, columns=['Correlation'])
correlation.dropna(inplace=True)

# print(correlation.sort_values('Correlation', ascending=False).head(10))
## Add Rating counts
correlation = correlation.join(ratings_average['ratings_count'])
# print(correlation.sort_values('Correlation', ascending=False).head(10))

##See the recommendations
recommendation = correlation[correlation['ratings_count']>100].sort_values('Correlation', ascending=False)
# print(recommendation.head(10))

##Confirm the recommendation Quality
recommendation = recommendation.merge(movies, on='title')
print(recommendation.head(10))

# print(ratings_average.sort_values(by = ['ratings_count', 'rating'], ascending = False))
```

Figure 2. Recommendation code

```
                 title  Correlation  ratings_count  movieId                                      genres
0       Matrix, The (1999)     1.000000            278     2571                     Action|Sci-Fi|Thriller
1          Die Hard (1988)     0.544466            145     1036                      Action|Crime|Thriller
2         Inception (2010)     0.514767            143    79132  Action|Crime|Drama|Mystery|Sci-Fi|Thriller|IMAX
3        Braveheart (1995)     0.496045            237      110                           Action|Drama|War
4            Aliens (1986)     0.470865            126     1200             Action|Adventure|Horror|Sci-Fi
5    Lion King, The (1994)     0.444932            172      364  Adventure|Animation|Children|Drama|Musical|IMAX
6   Monsters, Inc. (2001)     0.441205            132     4886    Adventure|Animation|Children|Comedy|Fantasy
7    Batman Begins (2005)     0.440338            116    33794                       Action|Crime|IMAX
8    Jurassic Park (1993)     0.427936            238      480          Action|Adventure|Sci-Fi|Thriller
9       Fight Club (1999)     0.417196            218     2959                Action|Crime|Drama|Thriller
PS D:\code>
```

Figure 3. Recommendation results

So I would recommend ‘Die hard’, or ‘Inception’ is also a good choice.