8/20/2015

Trinadh Gupta

DePaul University

ECT-584: Recommender Systems

Via Collaborative Filtering Techniques

Contents

[**ECT-584 Project - Recommender Systems through Collaborative Filtering (KNN, Weighted Average)** 2](#_Toc427795129)

[Background: 2](#_Toc427795130)

[Data Description: 2](#_Toc427795131)

[Project Summary: 3](#_Toc427795132)

[Code Description: 3](#_Toc427795133)

[1) eudistance(moviedb,user1,user2): 3](#_Toc427795134)

[2) pearsondistance(movidb,user1,user2): 3](#_Toc427795135)

[3) nearestneigbours(moviedb,person,n=5,similarity=pearsondistance): 3](#_Toc427795136)

[4) getRecommendations(moviedb,person,similarity=pearsondistance): 3](#_Toc427795137)

[5) getRecommendationsKNN(moviedb,person,similarity=pearsondistance): 3](#_Toc427795138)

[6) getRecommendationsKNNMAE(moviedb,person,similarity=pearsondistance): 3](#_Toc427795139)

[7) transformPrefs(moviedb): 3](#_Toc427795140)

[8) calculateSimilarItems(moviedb,n=10): 3](#_Toc427795141)

[9) calculateSimilarUsers(moviedb,n=3): 3](#_Toc427795142)

[10) getRecommendedItems(moviedb,itemMatch,user): 3](#_Toc427795143)

[11) loadmoviedb(path): 3](#_Toc427795144)

[Recommendations Via User based Collaborative Filtering: 4](#_Toc427795145)

[Recommendations Via Item Based Collaborative Filtering: 5](#_Toc427795146)

[Appendices-1: 7](#_Toc427795147)

[Source Code: 7](#_Toc427795148)

[Data Sets: 7](#_Toc427795149)

[Outputs: 7](#_Toc427795150)

[References 7](#_Toc427795151)

# **ECT-584 Project - Recommender Systems through Collaborative Filtering (KNN, Weighted Average)**

## Background:

One of the most challenging situations for most of the online business is to engage with the clients as much as possible and interact as much as possible to know more about their needs and convert the needs to business opportunities as quickly as possible, which means more business revenues and profits, at the same time making client happy with their purchase without throwing irrelevant stuff at them or making them navigate a whole lot of stuff before they arrive at what they want.

Recommender systems has been very effective in this aspect and has been the backbone for almost all big technology companies in the recent days such as Netflix, amazon, YouTube etc.

## Data Description:

GroupLens Research has collected and made available rating data sets from the MovieLens web site ([http://movielens.org](http://movielens.org/)). The data sets were collected over various periods of time, depending on the size of the set.

The data was collected through the MovieLens web site (movielens.umn.edu) during the seven-month period from September 19th, 1997 through April 22nd, 1998. This data has been cleaned up – users who had less than 20 ratings or did not have complete demographic information were removed from this data set.

This data set consists of:

\* 100,000 ratings (1-5) from 943 users on 1682 movies.

\* Each user has rated at least 20 movies.

Dataset format: each record in the data set contains

user id | item id | rating | timestamp.

**Note:** for the purpose of this recommender systems project we are only using file “u” in the zip file [ml-100k.zip](http://files.grouplens.org/datasets/movielens/ml-100k.zip)

## Project Summary:

The goal of the project is to build a recommender system for movie lens data set using collaborative filtering technique and is evaluated through mean absolute error.

The scripting language used for this is python.

## Code Description:

The code is broadly defined using 12 different functions listed below along with a detailed description

1. eudistance(moviedb,user1,user2): This function takes two users as input and calculates the Euclidian distance between them.
2. pearsondistance(movidb,user1,user2): This function takes two users as input and calculates the pearson distance between them.
3. nearestneigbours(moviedb,person,n=5,similarity=pearsondistance): This function takes a user as input and brings out the number of nearest neighbourse , based on the number you pass as an argument. By default it uses 5 as the number of neighbors to output. And the function uses Pearson correlation as a similarity measure.
4. getRecommendations(moviedb,person,similarity=pearsondistance): This function brings all movies that a user should watch, in a sorted order of user based similarities.
5. getRecommendationsKNN(moviedb,person,similarity=pearsondistance): This function brings all movies that a user hasn’t watch, but his/ her nearest neighbor has watched , in a sorted order of predicted ratings, this is calculated for k- nearest neighbors
6. getRecommendationsKNNMAE(moviedb,person,similarity=pearsondistance):This function evaluates the KNN by calculating the mean average error for the recommendations.
7. transformPrefs(moviedb): This function is used to invert the matrix from user based to movies based matrix.
8. calculateSimilarItems(moviedb,n=10): This function is used to create an item based similarity database, which can be useful for item based collaborative filtering. Based the number passed as an argument it generates the database
9. calculateSimilarUsers(moviedb,n=3): This function is used to create an user based similarity database, which can be useful for user based collaborative filtering. Based the number passed as an argument it generates the database.
10. getRecommendedItems(moviedb,itemMatch,user): This function is used to recommend items/ movies to a specific user using item based collaborative filtering technique.
11. loadmoviedb(path): The function takes the path as input from where it loads the data sets to support the recommendation system.

## Recommendations Via User based Collaborative Filtering:

1. User based collaborative filtering techniques to bring recommendations for a user. This has been achieved by two different approaches and please refer to the source code (in Appendices-1). For the purpose of the report I have uploaded results of the recommendations for the first 10 users. And I am providing the recommendations for User “4” below

Movies Saw by User 4:

{'In & Out (1997)': 5.0, 'Client, The (1994)': 3.0, 'Indiana Jones and the Last Crusade (1989)': 3.0, 'Spawn (1997)': 2.0, 'Incognito (1997)': 5.0, 'Blues Brothers 2000 (1998)': 5.0, 'Liar Liar (1997)': 5.0, 'Lost Highway (1997)': 5.0, 'Wonderland (1997)': 5.0, 'Mimic (1997)': 3.0, "One Flew Over the Cuckoo's Nest (1975)": 4.0, 'Starship Troopers (1997)': 4.0, 'Star Wars (1977)': 5.0, 'Desperate Measures (1998)': 5.0, 'Air Force One (1997)': 5.0, "Ulee's Gold (1997)": 5.0, 'Event Horizon (1997)': 4.0, 'Contact (1997)': 5.0, 'Cop Land (1997)': 5.0, 'Assignment, The (1997)': 5.0, 'Scream (1996)': 4.0, 'Seven (Se7en) (1995)': 4.0, 'Conspiracy Theory (1997)': 3.0, 'Wedding Singer, The (1998)': 5.0}

Recommendations for User 4:

* 1. Top 5 Recommended Movies to “User-4” via Weighted Average of all users

[(5.0, 'They Made Me a Criminal (1939)'), (5.0, 'Star Kid (1997)'), (5.0, "Someone Else's America (1995)"), (5.0, 'Sliding Doors (1998)'), (5.0, 'Saint of Fort Washington, The (1993)')]

*(Refer to txt file “totalweightedaveragepearson4.txt” in Appendice-1 below for detailed list of information)*

* 1. Weighted Average of K- Nearest Neighbors
     1. 3 Nearest Neighbors

[(5.0, 'Young Frankenstein (1974)'), (5.0, 'Wrong Trousers, The (1993)'), (5.0, 'When Harry Met Sally... (1989)'), (5.0, 'Wallace & Gromit: The Best of Aardman Animation (1996)'), (5.0, 'Unbearable Lightness of Being, The (1988)')]

*(Refer to txt file “resultknnpearson3.txt” in Appendice-1 below for detailed list of information)*

* + 1. 5 Nearest Neighbors

[(5.0, 'Young Frankenstein (1974)'), (5.0, 'Wrong Trousers, The (1993)'), (5.0, 'World of Apu, The (Apur Sansar) (1959)'), (5.0, 'Wild Bunch, The (1969)'), (5.0, 'When Harry Met Sally... (1989)')]

*(Refer to txt file “resultknnpearson4-5users.txt” in Appendice-1 below for detailed list of information)*

* + 1. 10 Nearest Neighbors

[(5.0, 'Wrong Trousers, The (1993)'), (5.0, 'World of Apu, The (Apur Sansar) (1959)'), (5.0, 'Winnie the Pooh and the Blustery Day (1968)'), (5.0, 'Wild Bunch, The (1969)'), (5.0, 'Wallace & Gromit: The Best of Aardman Animation (1996)')]

*(Refer to txt file “resultknnpearson4-10users.txt” in Appendices-1 below for detailed list of information)*

* 1. Mean Average Error Comparison K- Nearest Neighbors (3, 5, 10 Users)

Each User is described with number of movies he rated already. In the below chart

Eg: User1-271 means user 1 has already rated 271 movies and the mean average error for that user.

*(Refer to txt file “resultknnpearsoncurrent3.txt”,” resultknnpearsoncurrent4-5users.txt”,” resultknnpearsoncurrent4-10users.txt” in Appendice-1 below for detailed list of information)*

### User based Recommender Evaluation (Mean Average Error) – Comparison

The below chart provides a comparison of error variation in user based recommendation via mean average error for different sizes of nearest neighbors for first 5 users.

## Recommendations Via Item Based Collaborative Filtering:

1. Item based collaborative filtering techniques to bring recommendations for a user. This type of recommendation is very efficient as you don’t need to update the recommendations quiet frequently as you do the user based recommendation and aswell requires less computation effort over and all compared to user based recommendation. Also, this kind of recommendation system can be done during idle period of a system. This has been achieved by two different approaches and please refer to the source code (in Appendices-1). For the purpose of the report I have uploaded results of the recommendations for the first 10 users. And I am providing the recommendations for User “4” below

Movies Saw by User 4:

{'In & Out (1997)': 5.0, 'Client, The (1994)': 3.0, 'Indiana Jones and the Last Crusade (1989)': 3.0, 'Spawn (1997)': 2.0, 'Incognito (1997)': 5.0, 'Blues Brothers 2000 (1998)': 5.0, 'Liar Liar (1997)': 5.0, 'Lost Highway (1997)': 5.0, 'Wonderland (1997)': 5.0, 'Mimic (1997)': 3.0, "One Flew Over the Cuckoo's Nest (1975)": 4.0, 'Starship Troopers (1997)': 4.0, 'Star Wars (1977)': 5.0, 'Desperate Measures (1998)': 5.0, 'Air Force One (1997)': 5.0, "Ulee's Gold (1997)": 5.0, 'Event Horizon (1997)': 4.0, 'Contact (1997)': 5.0, 'Cop Land (1997)': 5.0, 'Assignment, The (1997)': 5.0, 'Scream (1996)': 4.0, 'Seven (Se7en) (1995)': 4.0, 'Conspiracy Theory (1997)': 3.0, 'Wedding Singer, The (1998)': 5.0}

Movies Recommended via Item based filtering With 10 Nearest Movies:

[(5.0, 'Young Frankenstein (1974)'), (5.0, 'Yankee Zulu (1994)'), (5.0, 'Wonderful, Horrible Life of Leni Riefenstahl, The (1993)'), (5.0, 'Wizard of Oz, The (1939)'), (5.0, 'Withnail and I (1987)'), (5.0, 'With Honors (1994)'), (5.0, 'Wishmaster (1997)'), (5.0, 'Winter Guest, The (1997)'), (5.0, 'Wild Reeds (1994)'), (5.0, 'Wife, The (1995)'), (5.0, 'Wedding Bell Blues (1996)'), (5.0, 'Walking Dead, The (1995)'), (5.0, 'Twilight (1998)'), (5.0, 'Turning, The (1992)'), (5.0, 'Stars Fell on Henrietta, The (1995)'), (5.0, 'Small Faces (1995)'), (4.833333333333333, "Wooden Man's Bride, The (Wu Kui) (1994)"), (4.666666666666667, 'Witness (1985)'), (4.5, 'Twin Town (1997)'), (4.4, 'Zeus and Roxanne (1997)'), (4.333333333333333, 'Á köldum klaka (Cold Fever) (1994)'), (4.0, 'Women, The (1939)'), (4.0, 'Window to Paris (1994)'), (4.0, 'Wild Things (1998)'), (4.0, 'Vermin (1998)'), (4.0, 'Venice/Venice (1992)'), (4.0, 'Unzipped (1995)'), (4.0, 'Underground (1995)'), (4.0, 'The Deadly Cure (1996)'), (3.6666666666666665, 'Wings of Courage (1995)'), (3.5, 'You So Crazy (1994)'), (3.0, 'War at Home, The (1996)'), (3.0, 'Unstrung Heroes (1995)'), (3.0, 'Trial by Jury (1994)'), (3.0, 'Tokyo Fist (1995)'), (2.0, 'War Room, The (1993)'), (2.0, 'Total Eclipse (1995)'), (2.0, 'That Darn Cat! (1965)')]

Movies Recommended via Item based filtering With 20 Nearest Movies (Few list):

[(5.0, 'Young Frankenstein (1974)'), (5.0, 'Yankee Zulu (1994)'), (5.0, 'Wonderful, Horrible Life of Leni Riefenstahl, The (1993)'), (5.0, 'Woman in Question, The (1950)'), (5.0, 'Wizard of Oz, The (1939)'), (5.0, 'Withnail and I (1987)'), (5.0, 'With Honors (1994)'), (5.0, 'Wishmaster (1997)'), (5.0, 'Winter Guest, The (1997)'), (5.0, 'Wings of Desire (1987)'), (5.0, 'Wild Bill (1995)'), (5.0, 'Wild America (1997)'), (5.0, "Widows' Peak (1994)"), (5.0, "White Man's Burden (1995)"), (5.0, 'White Balloon, The (1995)'), (5.0, 'When a Man Loves a Woman (1994)'), (5.0, 'When We Were Kings (1996)'), (5.0, 'When Harry Met Sally... (1989)'), (5.0, 'What Happened Was... (1994)'), (5.0, "Wes Craven's New Nightmare (1994)"), (5.0, "Wend Kuuni (God's Gift) (1982)"), (5.0, 'Welcome to the Dollhouse (1995)']

Movies Recommended via Item based filtering With 30 Nearest Movies (Few list):

[(5.0, 'Young Frankenstein (1974)'), (5.0, 'Yankee Zulu (1994)'), (5.0, 'Wonderful, Horrible Life of Leni Riefenstahl, The (1993)'), (5.0, 'Woman in Question, The (1950)'), (5.0, 'Wizard of Oz, The (1939)'), (5.0, 'Withnail and I (1987)'), (5.0, 'With Honors (1994)'), (5.0, 'Wishmaster (1997)'), (5.0, 'Winter Guest, The (1997)'), (5.0, 'Wings of Desire (1987)'), (5.0, 'Wild Bill (1995)'), (5.0, 'Wild America (1997)'), (5.0, "Widows' Peak (1994)"), (5.0, "White Man's Burden (1995)"), (5.0, 'White Balloon, The (1995)'), (5.0, 'When a Man Loves a Woman (1994)'), (5.0, 'When We Were Kings (1996)'), (5.0, 'When Harry Met Sally... (1989)'), (5.0, 'What Happened Was... (1994)'), (5.0, "Wes Craven's New Nightmare (1994)"), (5.0, "Wend Kuuni (God's Gift) (1982)"), (5.0, 'Welcome to the Dollhouse (1995)'), (5.0, 'Wedding Bell Blues (1996)'), (5.0, 'Warriors of Virtue (1997)'), (5.0, 'Walking Dead, The (1995)']

Movies Recommended via Item based filtering With 40 Nearest Movies (Few list):

[(5.0, 'Young Frankenstein (1974)'), (5.0, 'Yankee Zulu (1994)'), (5.0, 'Wonderful, Horrible Life of Leni Riefenstahl, The (1993)'), (5.0, 'Woman in Question, The (1950)'), (5.0, 'Wizard of Oz, The (1939)'), (5.0, 'Withnail and I (1987)'), (5.0, 'With Honors (1994)'), (5.0, 'Wishmaster (1997)'), (5.0, 'Winter Guest, The (1997)'), (5.0, 'Wings of Desire (1987)'), (5.0, 'Wild Bill (1995)'), (5.0, "Widows' Peak (1994)"), (5.0, "White Man's Burden (1995)"), (5.0, 'White Balloon, The (1995)'), (5.0, 'When a Man Loves a Woman (1994)'), (5.0, 'When We Were Kings (1996)')]

Movies Recommended via Item based filtering With 50 Nearest Movies (Few list):

[(5.0, 'unknown'), (5.0, 'Young Frankenstein (1974)'), (5.0, 'Yankee Zulu (1994)'), (5.0, 'Woman in Question, The (1950)'), (5.0, 'Wizard of Oz, The (1939)'), (5.0, 'Withnail and I (1987)'), (5.0, 'With Honors (1994)'), (5.0, 'Wishmaster (1997)'), (5.0, 'Winter Guest, The (1997)'), (5.0, 'Wings of Desire (1987)'), (5.0, "Widows' Peak (1994)"), (5.0, 'When a Man Loves a Woman (1994)'), (5.0, 'When We Were Kings (1996)'), (5.0, 'When Harry Met Sally... (1989)'), (5.0, 'What Happened Was... (1994)'), (5.0, "Wes Craven's New Nightmare (1994)")]

## Appendices-1:

### Source Code:

****

### Data Sets:

<http://grouplens.org/datasets/movielens/>



### Outputs:

****

****

****

****

## References

1. ECT-584 – Reference Material
2. Mining the Web: Transforming Customer Data into Customer Value , by George Linoff and Michael Berry
3. A Programmers guide to Data Mining : Ancient Art of the Numerati