Recommender System for Movies using Collaborative Filtering

Nan Yu Hlaing, May Phyo Oo
Computer University, Pathein
nan.nanyuhlaing87@gmial.com , mayphyooo@gmail.com

Abstract

Recommender systems are programs which attempt to predict items that a user may be interest in. Recommender systems acts personalized decision guides, aiding users in decisions on matter related to personal taste. Recommender system depends on information provided by different users to gather its knowledge. Collaborative Filtering (CF) become an important data mining technique to make personalized recommendations for books, web pages or movies, etc. One popular algorithm is the memory-based collaborative filtering, which predicts a user's preference based on his or her similarity to other users (instances) in the database. Movie recommendation system demonstrates the advantages of multidimensional visualization of the recommender system's results. In this paper, we implement a recommender system for movies using memory–based collaborative filtering method. Memory-based for collaborative filtering predicts average (weighted) rating between similar users or items.

Keyword: Recommender system, Collaborative filtering, Memory - Based Algorithms

1. Introduction

Recommender systems help individuals and communities address the challenges of information overload. Information filtering recommender looks at the syntactic and semantic content of items to determine which are likely to be of interest or value of user. Collaborative filtering recommenders use the opinions of other users to predict the value of items for each user in the community. Collaborative filtering aims at predicting the user interest for a given item based on a collection of user profiles.

Memory-based collaborative filtering measures similarities between test user and other user rating. . Memory-based CF algorithms retain all relevant data in memory and compute the required prediction on demand in real. Several algorithms have been proposed for memory-based CF due to their high accuracy and simplicity of implementation. Collaborative filtering is a three stage process of finding similar users, computing predicted rating and applying the predictions as recommendations to the user. In this paper, we present an approach to give the predictions result of user's requirements using rating by memory-based collaborative filtering.

Recommender system apply data analysis techniques to the problem of helping users find the items they would like to purchase at E-commerce site by producing a predicted likeness score or a list of Top N-list recommended items for a given active user. Recommender systems help overcome information overload by providing personalized suggestions based on a history of user's likes and dislikes. Many on-line stores provide recommend - ing services e.g. .Amazon, CDNOW, etc [2]. Recommendation systems are provided by the search engines or online vendors to the user. Recommendation systems are implemented in commercial and non- profit web sites to predict the user preferences .Recommendation system can generally be categorized in five major types [8]. They are Content-based filtering, Demographic filtering, Collaborative filtering; Knowledge-based Method and Utility-based Method. Recommendation system is widely used on the content-based filtering and collaborative filtering methods. In this paper, we used collaborative filtering method using Pearson Correlation Algorithm.

The rest of paper is organized as follows. Related works are discussed in section 2. Collaborative Filtering Method is explained in section 3. Section 4 describes memory based algorithm with reasons. Overview of the system is explained in section 5. Implementation of the system is presented in section 6. Finally we conclude this system in section 7.

2. Related Works

Nowadays, the most popular in movies recommender system is Movie Lens and use collaborative filtering movies [2, 8]. The more user ratings the user can give, the more accurate recommendations the user can get. Therefore its recommendation depends on large statistical data of user ratings.

Recommender.com includes a knowledge-based movie recommender system [3]. Knowledge-based recommender systems use knowledge about users and items to generate recommendations by reasoning about which items meet the user requirements [1]. But they do not attempt to build long term user model. In that system, user has to

enter a movie name to get recommendation. Then the system will recommend the movies which are similar to the movie the user just entered. But sometimes user has no idea to enter a specific movie name.

Another one of the amazon.com is movie recommendation website. Uses recommendations are like a targeted marketing tool and it can give recommendations on movie, music, book and many other products for each user. It uses Item-to-item collaborative filtering method to recommend items. Rather than matching the active user to similar users, item-to-item collaborative filtering matches each of the user's purchased and rated items to similar items, then combines those similar items into a recommendation list [2, 8].

3. Collaborative Filtering Method

Collaborative filtering techniques collect and establish profiles, and determine the relationships among the data according to similarity models. The possible categories of the data in the profiles include user preferences, user behavior pattern, or items properties. Collaborative filtering technology is being widely used on the web as an approach to information filtering and recommendation by commercial service providers like Amazon and Yahoo! The goal of collaborative filtering is to make recommendations for a test user by utilizing the rating information of users who share interests similar to the test user. Because rating is determined not only by user interest but also the rating habits of user, it is important to normalize rating of different users to the same scale.

Collaborative recommendations can be grouped into two general classes: memory-based (or heuristic-based) and model-based. Memory-based algorithms essentially are heuristics that make rate predictions based on the entire collection of previously rating items by the users. Model-based algorithms use the collection of rating to learn a model, which is then used to make rate predictions [4]. In our system, we use memory-based collaborative recommendation.

4. Using Memory-based Algorithms

A movie recommendation system is developed using memory-based algorithms. This algorithm can be summarized in the following steps:

1. Select same users that have items to rate similarity with active user and weight all users with respect to similarity with the active user. Similarity is measured as the Pearson correlation between active user and existing user rating.
2. Compute a prediction from a weighted combination of the selected existing rating.

Memory-based algorithm is the most popular prediction technique in CF applications [7]. The Pearson correlation is widely and successfully used as a similarity measure between users. Memory-based methods have the advantages of being able to rapidly incorporate the most up-to-date information and relative accurate prediction [7]. The new user has an average rated for movies he/ she has rated. Then the predicted rating of the new users over other products could be calculated by adding weighted sum of other users rating. The weights could be determined by the similarity between active user and other users [4, 7].

According to Pearson Correlation coefficient step1 can be proved that computing similarity between two users as shown in equation 1. After that, step 2 can also proved that calculating the predicted rating of active user for items j as shown in equation 2.

Step1: Compute the similarity between two users is computed using the Pearson correlation coefficient:

$$w(a,i) = \frac{\sum_j (r_{a,j} - \bar{r}_a)(r_{i,j} - \bar{r}_i)}{\sqrt{\sum_j (r_{a,j} - \bar{r}_a)^2 \sum_j (r_{i,j} - \bar{r}_i)^2}} \quad ---- (1)$$

Step 2: the predicted rating of new (active) user for items j:

$$P_{a,j} = \bar{r}_a + k \sum_{i=1}^{n} w(a,i)(r_{ij} - \bar{r}_i) \quad ---- (2)$$

$P a, j$ is a weighted sum of the rate of other user, n is the number of the user who rating product j, k is a normalizing factor. $\bar{r}_a$ is the average of active (new) user a's rating for all items. $w (a,i)$ is similarity between user a and i. $\bar{r}_i$ is average of existing user i's ratings for all items .$r_{i,j}$ is rating existing user a gave item.

5. Overview of the system

In our movie recommender system, have four main components: (i) get user rating for movie, (ii) calculate prediction using collaborative filtering, (iii) show top N movies title list with according to predicted rate value. (iv) save to database predicted top N movie list for active user. User can present rating to user interest's movies. Accept user rate from rate scale and rating scale have 1to 5 scale. For rating value and rate meaning are

1. I don't like.
2. It's ok.
3. I like it.
4. I love it.
5. I most like it.
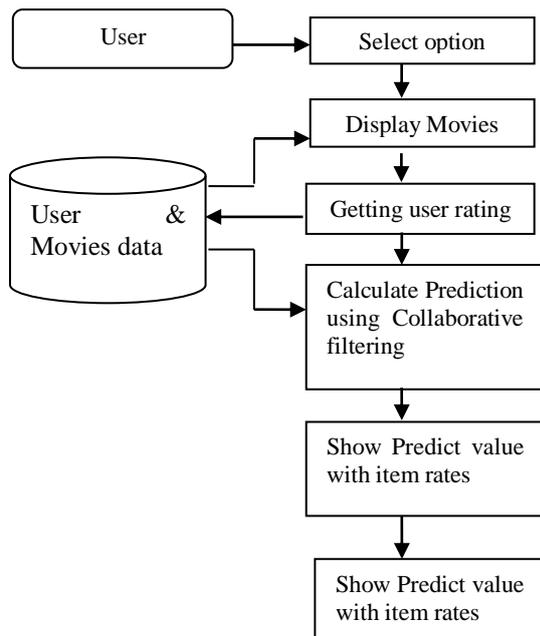
Figure1: Overview of the System Design

This rating is movie recommender system to use CF algorithms of memory-based approach. In our system, user can search by categories (genres of movies), by movies title, by director, by actor and by actress. In this system need user profile to give recommendations. We create user profile explicitly because explicit profiling can provide more confidence in the recommendation. This system new user can free register (sign up) because user can get recommendation and that information will be used in the process of computing recommendations.

If member user can choose select option for example by categories (genres), by movie title, by actor) may be like. The system accept query from user.

Step 1: User can rate for items or products like -minded. After got rate value, the system stores the dataset with item.

Table1. Average Rate Value for Movies

User Movies	mumu	susu	zawzaw	koko
Bruno	5	5	3	5
My sister keeper	2	3	2	2
The proposal	3	2	4	2
Average rate	3.3	3.3	3	3

This system computes weight (similarity) between active user and existing user using Pearson correlation algorithm .This algorithm is widely and successful similarity measure between users based on average rating. The results of equation (1) are described in table1.We suppose there are four users described in table. Mu Mu is active user. Su Su, Zaw Zaw and Ko Ko are old user. Their Similarity are calculated as follows:

W (mu mu, su su)   = o.8

W (mu mu, zaw zaw) = 0.3

W (mu mu, ko ko)   = 0.9

Step 2: After computing similarity between existing user and active user with equation (1). The equation (2) is used to predict recommendation of active user (mu mu).

Table2. Average Rate for Each Item

User Movies	mumu	susu	zaw zaw	koko
Bruno	5	5	3	5
My sister Keeper	2	3	2	2
The proposal	3	2	4	2
Bride War	?	4	1	3
Average rate	3.3	3.2	2.5	2.7

P (mumu, bride war) =3.3+1[0.64-0.45] = 3.49

According to table 2, predict rate all items are computed using equation (2) and show top-N movie list recommendation for active user. Separated four quarter for year can be seen in table 3 and user can search by years.

Table3. Quarter for Year

Quarter name	From Date	To Date
1st qtr	01-Jan-	31-Mar-
2nd qtr	01-Apr-	30-Jun-
3rd qtr	01-Jul-	30-Sep-
4th qtr	01-Oct-	31-Dec-

In this system, users can search by movies categories according to table 4.

Table4. Type of Categories

No	Categories
1	Comedy
2	Romance
3	Drama
4	Horror
5	Thriller
6	Action
7	Fantasy
8	Adventure
9	Crime
10	Sci-Fi

Then the system computes the prediction rate for the new (active) user and displays top N movies list result to the active user for nearest of he/she interested. After show result of top N- movie list for active user, this result saves into database.

6.   Implementation of the System

In this section, implementation of this system can be seen in following.

Figure 2: Movies Page for Rate

Memory-based algorithms utilize the entire user-item database to generate a prediction. Memory-based methods have the advantages of being able to rapidly incorporate the most up-to-date information and relative accurate predict recommendation. Hence, user can easily search the movies they want to rate due to the advantage of memory method. Moreover, users can also rate movies from new movies, old movies.

The active user filters out of search movies list by user like-minded. The user can rate to movie in this list and this save rate value and then compute prediction show in figure 2.
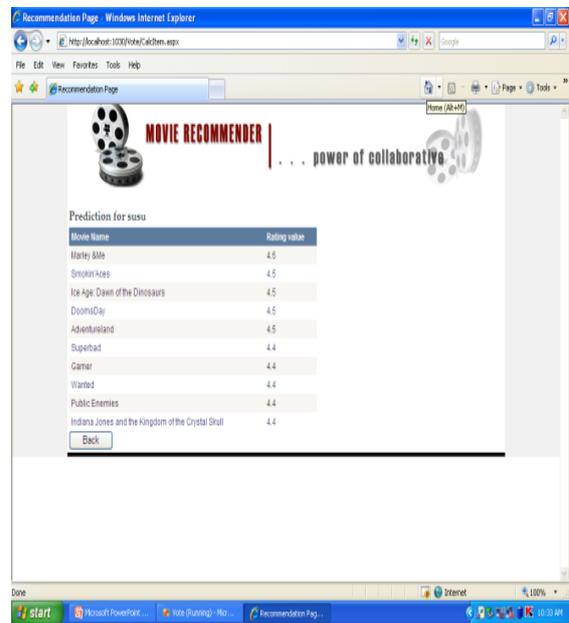
Figure3:   Predict Recommendation Page for Active   User

After getting rating from user, this system computes predicted recommendation for active user. And then this system show result of top –movie list according to rating value show figure3.

In this way, this system can give predict item recommendations to users.

6. Conclusion

The recommendation systems can predict user behavior patterns without any knowledge of the user in advance, and to evaluate the accuracy by the comparing the prediction and the reality. In our system, user can get detail movies information of the recommend movies from the system. Our system use memory based algorithm reduces time and get extensive information. This algorithm used not only movies website but also another internet application such as laptop, music, book store, web page, etc. User can search easily genres, title, actor, actress, and year. This system is to present user with highly relevant set of items.

7. References

[1] G.Adomavicius and A.Tuzhilin, "*Toward the next generation of recommender system: a*

survey of the state-of –the art and possible extentions" IEEE Transaction on Knowledge and data Engineering 17(6):734-749

[2] J.S.Breese, D.Heckerman, C. Kadie *"Empirical Analysis of Predictive Algorithms or Collaborative filtering",* proceeding of the 14th Conference on Uncertainty in Artificial Intelligence, 1998.

[3] R. Burke, *"Knowledge-based Recommender Systems".* In: A. Kent, (ed.): Encyclopedia of Library and Information Systems, vol. 69, Supplement 32.

[4] G.linden, B.Smith and J.York,"*Amazon.com Recommendation: Item-to-Item Collaborative filtering",* Industry Report from IEEE Internet Computing.

[5] P. Melville and J.Mooney Raymond and Ramadass Nagaraja *"Content-Boosted Collaborative filtering for Improved Recommendation"* Department of Computer Sciences University of Texas Austin, TX 78712

[6] R.Mukherjee, P.s.Dutta and S.Sen," Movies2Go-*A new approach to online movies recommendation "*In Proceeding of the IJCAI Workshop on Intelligent Techniques for Web Personalization.

[7] Y. Qu, X. Yang, T. Huang" *Survey of Recommendation Systems and Algori- thms*" EE380L: Data mining.

[8]M.M .Yee," *Content-based movie recommender system*" University of Computer Studies, Yangon, Myanmar