Memory-Based Personalized Movie Recommender System

Khaing Mar lar Myint
Computer University (Mandalay)
Khaingmarlarmyint09@gmail.com,myintmyintmaw09@gmail.com

Abstract

The World Wide Web information grows explosively in the Internet and people encounter problem to pick some correct things from the overwhelming set of choices. The recommender systems help them choose something they actually want or need. Therefore, the recommender systems get the vital role in the Internet. One of the most successful technologies for recommender systems is called collaborative filtering. In this movie recommender system use memory-based collaborative filtering and firstly, the user has to give rating what he likes in the movie he has seen. It depends on that rating to find out the neighbors of an active user by using Pearson correlation coefficient method. To predict the movies that the active user has not seen depends on neighbors of the active user and the highest rating will be shown as recommended movie list.

1. Introduction

The amount of information on the Web is increasing according to the growth of information and communication infrastructure. As a result, recommender systems (RSs) for personalization are required [2].Recommender systems are being by an ever-increasing number of E-commerce sites to help consumers find products to purchase [3]. E-commerce recommendation algorithms often operate in a challenging environment, especially for large online shopping companies like eBay and Amazon. Usually, a recommender system providing fast and accurate recommendations will attract the interest of customers and bring benefits to companies [8]. One of the most popular and successful techniques that has been in recommender systems is known as collaborative filtering. In this system, collaborative filtering compares users according to their preference .Therefore, a database of users preferences must be need .The preference can be collected explicitly. In the first case the user's participation is required .The user explicitly submits his/her rating of movie on a rating scale from 1to 5.Based on this rating scale, compute similarity using Pearson Correlation Coefficient and predict movies for active user that he might like.

The rest of the paper is organized as follows: Section 2 is the related work .Section 3 explains the theory used in this system. In section 4, proposed system design is described and section5 evaluates experimental result of the prediction quality .Finally, conclusion is expressed in section 6.

2. Related works

Amazon.com uses recommendations as 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 [3].

Collaborative filtering was first introduced in the Tapestry email sytem.Trapestry was a very simple method yet effective as it introduced collaborative filtering [1].

In content–based recommendation systems recommend an item to user based upon a description of the item and a profile of the user's interest. While a user profile may be entered by the user, it is commonly learned from feedback the user provides on items. A variety of learning algorithms have been adapted to learning user profiles, and the choice of learning algorithm depends upon the representation of content [6].

3. Theoretical background

In this section, this paper explains the background theories used to implement personalized movie recommender system.

3.1 Personalized recommender system

When the World Wide Web becomes an increasingly popular medium, information overload intensifies: users are overwhelmed by which information should be consumed. Fortunately, recommender system offers a feasible solution to this issue. However, since the volume of transactions and web activity are increasing, it is not trivial to make useful recommendation. As such, how to select candidate items for personalized recommendation becomes critical. Personalization methods are often classified into broad categories, according to their recommendation approach and algorithmic techniques. Content-based filtering and collaborative filtering and rule base filtering are common methods to provide personalized recommendations. [7].

3.2 Collaborative filtering systems

The task in collaborative filtering is to predict the preference of an active consumer to a given product based on a database of consumers' product preferences. There are two kinds of collaborative filtering. They are: memory–based collaborative filtering and model-based collaborative filtering.

3.2.1 Model-based collaborative filtering

Model-based collaborative filtering, in contrast, uses the consumers' preference database to learn a model, which is then used for predications. The model can be built off-line over several hours or days. Model-based methods may prove practical for environments in which consumers' preferences change slowly with respect to the time needed to build the model. Model-based methods, however, are not suitable for environments in which consumers' preference models must be updated rapidly or frequently [4].

3.2.2 Memory-Based Collaborative Filtering

Memory-based *CF* algorithms use the entire or a sample of the user-item database to generate a prediction. Every user is part of a group of people with similar interests. By identifying the so-called neighbors of a new user (or active user), a prediction of preferences on new items for him or her can be produced.

Similarity computation between users is a critical step in memory-based collaborative filtering algorithms. There are many different methods to compute similarity or weights between users. The most commonly used similarity computation methods are

- Cosine-based similarity
- Pearson -correlation-based similarity
- Adjusted-cosine similarity

Since Pearson -correlation approach outperform the Cosine –based approach [7]. In this system use Pearson correlation based similarity.

In memory-based Pearson-correlation compute similarity between users a and b is given by

$$W(a,b)=\frac{\sum_{K}(V_{a,k}-\overline{V}a)(V_{b,k}-\overline{V}b)}{\sqrt{\sum_{K}(V_{a,k}-\overline{V}a)^{2}(V_{b,k}-\overline{V}b)^{2}}} \qquad (1)$$

Where, $V_{a,k}$ is the rating by user a on item k, $\overline{V}_a$ is the mean rating by user a and K is the set of items co-rated by both user a and b.

W (a, b) must generate a similarity score. It ranges from -1 (a perfect negative relationship) to +1 (a perfect positive relationship), with 0 stating that there is no relationship whatsoever.

To obtain predictions or recommendations is the most important step in a collaborative filtering system. In the user-based *CF* algorithm, subsets of nearest neighbors of the active user are chosen based on their similarity with him or her, and a weighted aggregate of their ratings is used to generate predictions for the active user [4].

To make a prediction for the active user, *a*, on a certain item j, we can take a weighted average of all the ratings on that item according to the following formula.

$$Pa,j=\overline{V}a+\frac{\sum_{B}W(a,b)(V_{b,j}-\overline{V}_{b})}{\sum_{B}W(a,b)} \qquad (2)$$

Where $\overline{V}_a$ and $\overline{V}_b$ are the average ratings for the user a and user b on all other rated items, and W (a, b) is the similarity between the user a and user b. B is the set of neighbors being considered.

4. Proposed system design

This system uses the memory-based collaborative filtering: Pearson correlation similarity to implement purposed system.

4.1. Memory-based collaborative filtering

The memory-based collaborative filtering method makes recommendations according to the following simple step by step procedure:

1. Users are requested to give preference ratings to movies.
2. A recommender system uses the ratings in order to determine which users' ratings are the most similar to active user's ratings.

3. The system predicts ratings of movies for the active user have not seen, based on the ratings of similar users.
4. According to the rating of predicted movies, select movies that have the highest rating to compose the recommendation set and recommend them to active user.

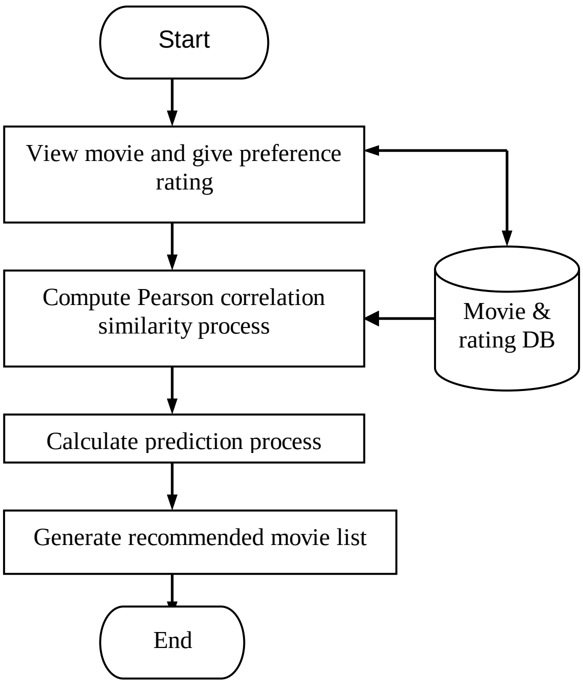

Figure 1. Overview of system design

4.2 Case study of proposed system

In the first step, user gives the rating to movies he has seen among those shown by the system. Rating values are from 1 star to 5 stars .5 stars means “must seen”, 4 stars means “will enjoy”, 3 stars means “It is ok”, 2 stars is “fairly bad”, and 1star means “Awful”. This system stores these rating values to the database as shown in table 1. In table1, Avg means average value with respect to active user.

Table 1.Sample of user-movie rating matrix

	M1	M2	M3	M4	M5	Avg
AyeAye	4	3	2	5	5	3.5
Ni Ni	3	4	1	4	4	3.5
MoeThu	4	1	2	4	4	2.5
HtetHtet	1	2	5	1		1.5
LinLin	4	2	1		4	3
Active user	4	3				3.5

In the second step, active user’s neighbor will be found by using Pearson Correlation Coefficient method. Pearson Correlation Coefficient method is mentioned in equation (1) of section (3.2.2).Pearson Correlation method will find active user’s neighbors depending on previous user. Find similarity between Active user and Aye Aye is

W (AyeAye, Active user) =

$$\frac{(4-3.5)(4-3.5)+(3-3.5)(3-3.5)}{\sqrt{(4-3.5)^2+(3-3.5)^2}\sqrt{(4-3.5)^2+(3-3.5)^2}}$$

=1

The similarity between active user and other users are calculated in the similar manner.
W (NiNi, Active user) = -1
W (MoeThu, Active user) = 1
W (HtetHtet, Active user) = -1
W (LinLin, Active user) = 1

The result of similarity value ‘0’ means no correlation -1 means dissimilarity and 0<similarity<=1 means two users are similar.So, neighbors of active user are Aye Aye, MoeThu and LinLin.

In the third step, calculate prediction is based on neighbors of active user. System predicts the movies that active user has not seen will be solved by using the equation (2) of section (3.2.2).Calculated prediction for Movie3, Movie4, and Movie5 are shown below.

P(Activeuser,M3)= $3.5+\frac{(2-3.5)1+(2-2.5)1+(1-3)1}{1+1+1}$

=2.167

P (Active user, M4) = $3.5+\frac{(5-3.5)1+(4-2.5)1}{1+1}$

=5

P(Active user, M5) = $3.5+\frac{(5-3.5)1+(4-2.5)1+(4-3)1}{1+1+1}$

=4.833

In step four, according to the predicted rating of movies, system selects top N movies that have highest set and recommend them to active user. Recommended movie page is shown in Figure2.

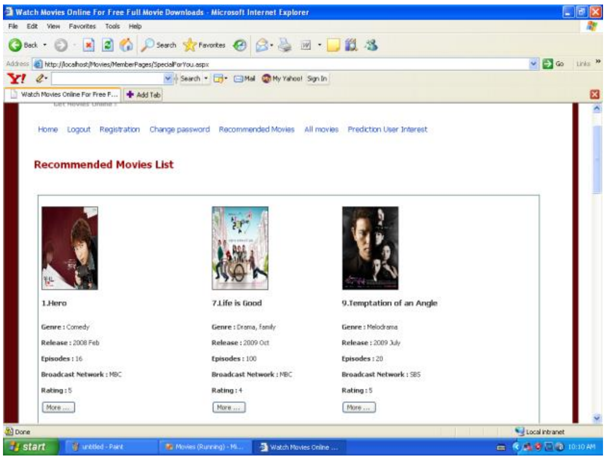

Figure 2.Recommendation page

5. Experiment result

The measurement method of evaluating the recommendation quality of recommendation system mainly includes statistical precision measurement method and decision supporting precision measurement method.

Statistical precision measurement method adopts MAE (Mean Absolute Error) to measure the recommendation quality .MAE is a commonly used recommendation quality measurement method. In this system, use MAE as the measurement criteria, which compute the average of absolute different between the prediction and true ratings

$$\text{MAE} = \frac{\sum_{(i,j)} | p_{i,j} - r_{i,j} |}{n} \qquad (3)$$

where *n* is the total number of ratings over all users, *pi,j* is the predicted rating for user *i* on item *j*, and *ri,j* is the actual rating. The lower the *MAE*, the better the prediction [5].Increase the number of neighbors; the experiment result is show in figure.3.

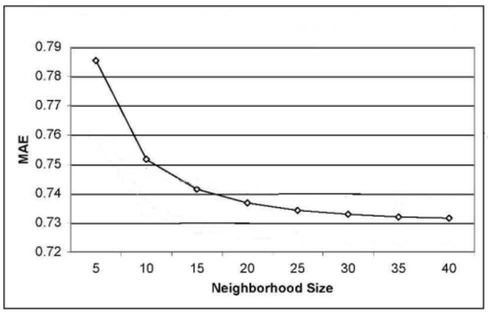

Figure 3.Rrecommendation quality (MAE) with respect to neighborhood size

6. Conclusion

In this system, present the recommendation systems are more and more needed because of huge amount of information on the internet. This movie recommender system predicts users' fondness of movies they have not seen. They do it by just rating the movies they have seen and depending on neighbor users who have similar fondness. By predicting like that, the users can differentiate the movies they have not seen whether it is good or bad without using a lot of time and without navigating difficultly. To gain more exact prediction accuracy MAE, the system solves with the condition that depends on neighbors of user. According to experiment result, the greater the numbers of neighbor users, the more exact the prediction accuracy.

7. References

[1] A. Sciberras , "Movie Collaborative Filtering",Dept Artificial Intelligence, University of Malta.

[2] H.J. Kwon, T.H. Lee, and K.S. Hong , "Improved Memory-based Collaborative Filtering Using Entropy-based Similarity Measures", School of Information and Communication Engineering , Sungkyunkwan University.

[3] J.B. Schafer, J. Konstan, and J. Riedl, "Recommender Systems in E-commerce", GroupLens Research Project, Department of Computer Science and Engineering, University of Minnesota Minneapolis.

[4]K. Yu, Z. Wen ,X.Xu , and M. Ester, "Feature Weighting and Instance Selection for Collaborative Filtering", Proc. 2nd International Workshop on Management of Information on the Web-web Data and Text Mining (MIW'01).

[5] L. Zheng, Y. Wang, J. Qi, and D. Liu, "Research and Improvement of Personalized Recommendation Algorithm Based on Collaborative Filtering",IJCSNS international Journal of Computer Science and Network Security , VOL 7, No.7, July 2007 .

[6] M.J. Pazzan , and D. Billsus, "Content –based Recommendation Systems", Rutgers University.

[7] T.Y. Tang, P. Winoto , and K.C.C. Chan, "Scaling Down Candidate Sets Based on the Temporal Feature of Items for Improved Hybrid Recommendations", Department of Computing, Hong Kong Polytechnic University ,Department of Computer Science , University of Saskatchwan,Canda .

[8] X. Su, and T.M. Khoshgoftaar, "A Survey of Collaborative filtering system".