International Journal of Business Analytics and Intelligence
7 (2) 2019, 42-46
http://publishingindia.com/ijbai/

Submitted: 29 April, 2019
Revised: 21 May, 2019
Accepted: 29 May, 2019

Recommending Movies Using User-Based and Item-Based Collaborative Filtering on R Platform

Chirag Malik*

Abstract

Recommender systems are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of a large amount of dynamically generated information according to user's preferences, interest or observed behaviour about an item. Recommender system has the ability to predict whether a particular user would prefer an item or not based on the user's profile. Recommendation algorithms are best known for their use on e-commerce websites, where they use inputs about a customer's interests to generate a list of recommended items. A new emerging sector in India, online movie viewership and subscription of movies, demands an expert and highly technical skill to understand the movie preferences of the viewers spread across India. This paper uses two techniques (run on R platform), user-based collaborative filtering and item-based collaborative filtering, to understand the preferences of people (without giving any reason to it) and recommending mechanism was solely based upon user-user similarity matrix and item-item similarity matrix. A dataset of 563 movies and 9,985 consumers from Amazon prime has been taken for recommending movies for people who have not watched a particular set of movies. The robustness of the two techniques is also compared and explained.

Keywords: User-Based Collaborative Filtering, Item-Based Collaborative Filtering, Recommendation Engine

Introduction

The explosive growth in the amount of available digital information and the number of visitors to the internet have created a potential challenge of information overload which hinders timely access to items of interest on the internet. Information retrieval systems, such as Google, DevilFinder and Altavista, have partially solved this problem but prioritization and personalization (where a system maps available content to user's interests and preferences) of information were absent. This has increased the demand for recommender systems more than ever before. Recommender systems are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of large amount of dynamically generated information according to user's preferences, interest or observed behaviour about the item. Recommender system has the ability to predict whether a particular user would prefer an item or not based on the user's profile.

Recommender systems are beneficial to both service providers and users. They reduce transaction costs of finding and selecting items in an online shopping environment. Recommendation systems have also proved to improve the decision-making process and quality. In e-commerce setting, recommender systems increase revenues, for the fact that they are effective means of selling more products. In scientific libraries, recommender systems support users by allowing them to move beyond catalogue searches. Therefore, the need to use efficient and accurate recommendation techniques within a system that will provide relevant and dependable recommendations for users cannot be over-emphasised.

Recommendation algorithms are best known for their use on e-commerce websites, where they use input about a customer's interests to generate a list of recommended items. A new emerging sector in India, online movie viewership and subscription of movies demand an expert and highly technical skill to understand the movie preferences of the viewers spread across India. This

* Associate Professor, BML Munjal University, Gurgaon, Haryana, India. Email: chiragmalik@yahoo.com

paper uses two techniques (run on R platform), user-based collaborative filtering and item-based collaborative filtering, to understand the preferences of people (without giving any reason to it) and recommending mechanism was solely based upon user-user similarity matrix and item-item similarity matrix. A dataset of 563 movies and 9,985 consumers from Amazon prime has been taken for recommending movies for people who have not watched a particular set of movies. The robustness of the two techniques is also compared and explained.

Literature Review and Research Problem

Previous personalised movie recommendation system focus only on viewers' historical viewing records or demographic data. This study proposes a new recommending mechanism from a user-oriented perspective. New recommender system, which is user-based collaborative filtering (UBCF) and item-based collaborative filtering (IBCF), is free from AIMED (Activities, Interests, Moods, Experiences, and Demographic information) characteristics. This paper aims at using R platform (free resource available for corporate and consultants) for recommending movies on online movie subscription websites using UBCF and IBCF.

Digital TV allows people to access numerous and varying kinds of TV programmes without space-time constraints. Although electronic programme guides (EPGs) can increase the accessibility of TV programmes, they often overload users by providing too many programme options. Therefore, the usefulness of EPG is an important issue in DTV design. To enhance the usefulness of EPGs, researchers have begun to develop TV recommendation mechanisms (i.e., embedded assistance) to help users access the TV programmes of their choice. Recommendation mechanisms operate by collecting viewing history data over a period of time, gleaning user viewing preferences from the data, mapping user preferences for TV programme attributes, filtering non-interesting programmes, and finally recommending an appropriate programme (Frank et al., 1980). In general, these TV programme recommendation mechanisms can be divided into two types-content-based filtering and collaborative filtering. Content-based filtering recommendation operates by automatically tracking each user's TV viewing patterns. Programme descriptors, such as the programme category, the name of the actor or actress, programme keywords, viewing time period, and so on, and a set of similarity metrics of the user profile are collected. Using these data, the recommender builds a user profile that represents the viewing preferences of

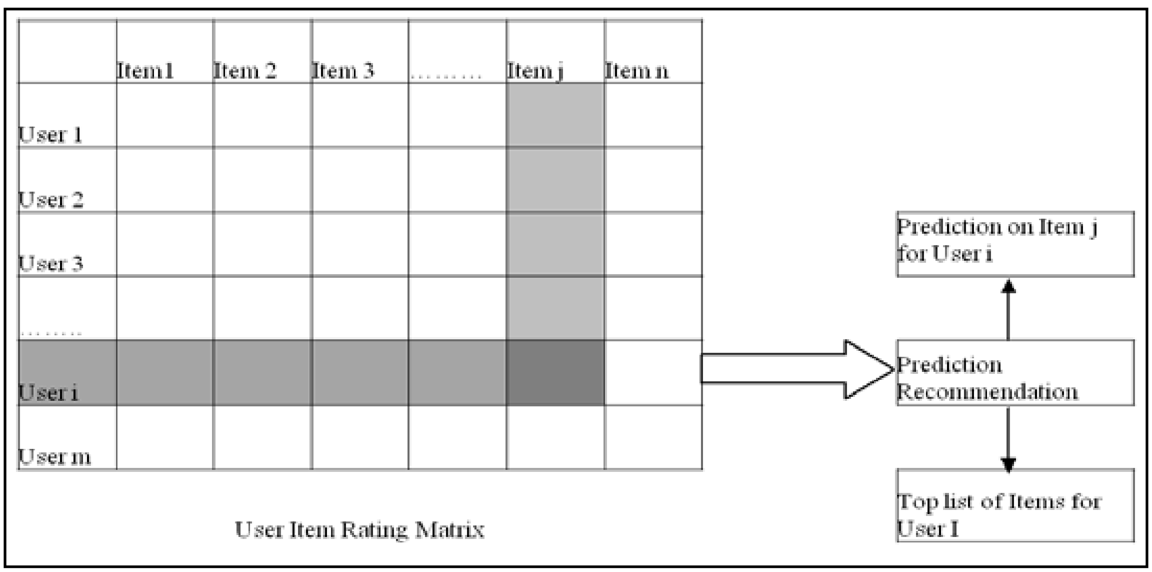

Fig. 1: Collaborative Filtering Process

each

The recommendation system stores massive amounts of viewer preference data in a database. The system then recommends programmes to users based on programme ratings given by people with similar profiles (Ansari et al., 2000; Ang, 1996). This system effectively forms a viewing community whose members share similar viewing preferences or similar viewing habits. In this virtual community, viewers can share information or recommend programmes to other viewers. Examples of this kind of recommender include the TV-scout system developed the peer-to-peer (p2p) recommender system proposed by Wang et al. (2006). Although both types of recommenders can assist TV viewers with programme selection, they both have some weaknesses. For example, a content-based filtering recommender only follows data from users' viewing histories and their viewing behaviour. Therefore, it cannot extend to other types of programmes or provide new programmes to viewers. This problem causes over-specialization in recommended programmes. On the other hand, a collaborative filtering recommender depends on other viewers' suggestion data, without which the mechanism may not make good recommendations, i.e., AIMED - Activity, Interest, Mood, Experience, and Demographic information of the user (Hsu et al., 2006). The AIMED recommender is a hybrid recommendation system based on the content-based and collaborative filtering methods. It makes programme suggestions using not only viewers' personal profile prediction criteria such as demographics, lifestyle and explicit preference information, from conventional recommendation models, but also inputs users' viewing contextual information, such as mood and viewing behaviour, into a prediction model (Burke, 2002; Alspector, 1998). By doing so, it has two advantages. Firstly, the recommender can avoid the weaknesses of conventional content-based filtering and collaborative filtering while taking advantage of their strengths. At the initial use, programme recommendation can be inferred from the viewer group's preference when the information of the user's viewing history is not available yet (Gena et al., 2001). As the system gathers more information about the user's viewing context, adding viewing context into the prediction of programme preference can fine-tune programme recommendation to match the personal preference of an individual user (Basu et al., 1998; Perse, 1998). Secondly, a well-trained AIMED recommender could suggest a suitable programme to users by considering both long-term programme preferences and the particular viewing context. Therefore, an AIMED recommender is able to adapt to the viewing context (Balabanovic et al., 1997). But, given the data of AIMED characteristics, the process of running the data and coming out with recommendation is a time-consuming and complex process. Another way to process the data only on the basis of item similarity matrix and user similarity matrix. This method doesn't demand any prior information of AIMED and solely works on the similarity matrix mechanism.

Two most important similarity measures are correlation-based and cosine-based. Pearson correlation coefficient is used to measure the extent to which two variables linearly relate with each other and is defined as:

$$sim(a,b) = \frac{\sum_{p \in P}(r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P}(r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P}(r_{b,p} - \bar{r}_b)^2}} \qquad \ldots(1)$$

where;

a, b; users

ra, p: rating of user a for item p

set of items, rated by both a and b possible similarity values between -1 to 1.

The Common Prediction Functions

$$pred(a,p) = \overline{r_a} + \frac{\sum_{b \in N} sim(a,b) * (r_{b,p} - \overline{r_b})}{\sum_{b \in N} sim(a,b)} \qquad \ldots(2)$$

The second method is the cosine similarity measure. It produces better results in item-based collaborative filtering. In cosine similarity measures, ratings are seen as vectors in n-dimensional space, and the similarity is calculated based on the angle between the vectors.

$$sim(\vec{a},\vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|} \qquad \ldots(3)$$

To compare the recommendation outcome of UBCF method and IBCF method, the deviation between predicted rating and actual ratings (also known as prediction accuracy) have been calculated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - r_i| \quad \ldots(4)$$

and

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - r_i)^2} \quad \ldots(5)$$

Objective of the Research

To run the movie recommendation engine using recommendation algorithm on R and recommend movies for the active and non-active users and compare the robustness of user-based collaborative filtering (UBCF) and item-based collaborative filtering (IBCF).

Research Methodology

A dataset of 563 movies and 9,985 consumers from Amazon prime has been taken for recommending movies for people who have not watched a particular set of movies. The sample dataset was taken with the due approval and permission of the Amazon India authorities and the recommender engine using R was applied. The techniques used were user-based collaborative filtering (UBCF) and item-based collaborative filtering (IBCF). To measure the robustness of the method, a separate set of data was used and a classification table was made.

Findings of the Research

It has been found the user-based recommendation was more robust as it has less error (Root Mean Square Error (RMSE), Mean Square Error (MSE) and Mean Absolute Error (MEA) or more accuracy percentage in the classification table. However, both methods are equally important to recommend movies to users.

Table 1: Comparative Values of RMSE, MSE and MAE for UBCF and IBCF Methods

	RMSE	*MSE*	*MAE*
UBCF	0.341864916	0.116871621	0.237863157
IBCF	0.476725497	0.2272672	0.255404009

Table 2: Top 10 Scores Using UBCF and IBCF

TOP 10 Score User Based Rec. Filtering		*TOP 10 Score Item Based Rec. Filtering*	
The Situation Room with Wolf Blitzer	0.111882811	The Situation Room with Wolf Blitzer	0.064647966
NBC Nightly News	0.077134284	NBC Nightly News	0.267689573
Dancing with the Stars	0.112713567	Dancing with the Stars	0.333699171
The Colbert Report	0.077134284	The Colbert Report	0.017279588
Larry King Live	0.077134284	Larry King Live	0.061431732
Everybody Loves Raymond	0.153878714	Everybody Loves Raymond	0.305660932
NHL Hockey	0.113013825	NHL Hockey	0.223449589
Campbell Brown: No Bias, No Bull	0.077134284	Campbell Brown: No Bias, No Bull	0.172957837
NBA Basketball	0.077134284	NBA Basketball	0.184547026
Two and a Half Men	0.077134284	Two and a Half Men	0.298187223

Table 3: Top 10 Recommendations Using UBCF and IBCF

Sr. No.	TOP 10 Recommendation Based Upon UBCF	TOP 10 Recommendation Based Upon IBCF
1	Everybody Loves Raymond	Dancing with the Stars
2	NHL Hockey	Everybody Loves Raymond
3	Dancing with the Stars	Two and a Half Men
4	The Situation Room with Wolf Blitzer	NBC Nightly News
5	NBC Nightly News	NHL Hockey
6	The Colbert Report	NBA Basketball
7	Larry King Live	Campbell Brown: No Bias, No Bull
8	Campbell Brown: No Bias, No Bull	The Situation Room with Wolf Blitzer
9	NBA Basketball	Larry King Live
10	Two and a Half Men	The Colbert Report

Implications of the Study

As online shopping and movies subscription is an emerging sector in India, it has become all the more important to understand the subscription behaviour and purchase patterns of the consumers. This method will enable the marketers to understand the purchase patterns of online consumers and recommend for non-buyers and existing buyers. This method will also help match the user and item associations.

References

Alspector, J., Kolez, A., & Karunanithi, N. (1998). *Comparing feature-based and clique-based user models for movie selection.* In Proceedings of the Third ACM Conference on Digital Libraries, Pittsburgh PA, ACM Press, New York.

Ang, I. (1996). *Living room wars.* Routledge, London.

Ansari, A., Essegaier, S., & Kohli, R. (2000). Internet recommendation systems. *Journal of Marketing Research, 37*(3), 363–375.

Ardissono, L., Gena, C., Torasso, P., Bellifemine, F., Difino, A., & Negro, B. (2004). User modeling and recommendation techniques for personalized electronic program guides. In L. Ardissono, A. Kobsa & M. Maybury (Eds.), *Personalized Digital Television: Targeting Programs to Individual Viewers* (pp. 3–26). Kluwer Academic Publishers, Dordrecht.

Balabanovic, M., & Shoham, Y. (1997). FAB: Content-based collaborative recommender. *Communications of the ACM, 40*(3), 66–72.

Basu, C., Hirsh, H., & Cohen, W. (1998). Recommendation as classification: Using social and content-based information in recommendation. In *Recommender System Workshop, 98,* 11–15.

Baudisch, P., Brueckner, L., & Scout, T. V. (2002). *Guiding Users from Printed TV Program Guides to Personalized TV Recommendation.* In Proceedings of the 2nd Workshop on Personalization in Future TV, (pp. 157–166). Malaga, Spain.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User Adapted Interaction, 12*(4), 331–370.

Frank, R., & Greenberg, M. (1980). *The public's use of television: Who watches what and why.* Sage, Beverly Hills.

Gena, C., & Ardissono, L. (2001). On the Construction of TV Viewer Stereotypes Starting from Lifestyle Surveys. In Workshop on Personalization in Future TV, Sonthofen, Germany.

Hsu, H. S., Wen, M. H., & Lee, C. H. (2006). *An Activity-Oriented Approach to Designing a User Interface for Digital Television.* In Proceedings of the 4th Euro iTV Conference, Athens, Greece, (pp. 83–90).

Wang, J., Polwelse, J., Fokker, J., & Reinders, M. J. T. (2006). *Personalization of a Peer-to-Peer Television System.* In Proceedings of the 4th EuroiTV Conference, Athens, Greece, (pp. 147–155).