IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-4 E-ISSN: 2347-2693

Addressing Cold Start Problem in Recommendation Systems with Collaborative filtering and Reverse Collaborative Filtering

Saniya Zahoor

NIT, Srinagar, India

Corresponding Author: saniya.zahoor@yahoo.com

Available online at: www.ijcseonline.org

Received: 13/Mar/2018, Revised: 21/Mar/2018, Accepted: 04/Apr/2018, Published: 30/Apr/2018

Abstract— Today Recommender system predicts the future preferences of the user based on the user's profile. A number of approaches have been taken to address the issue of recommendations, be it user based filtering methods, item-based filtering methods etc. The popular is Collaborative filtering technique used by some renowned companies like Amazon, YouTube and others. But the problem that still holds is the cold start problem and the amount of time and accuracy that is associated with these algorithms. A recent improvement suggested is the Reverse Collaborative filtering for the accuracy and pre-processing time. This paper implements and compares collaborative and reverse collaborative filtering solutions to address the cold start problem.

Keywords— Personalization, Profiles, Recommendation Systems, Cold Start Problem

I. INTRODUCTION

The explosive growth in the internet users has created a potential challenge of information overload which hinders timely access to data of interest on the Internet. Information retrieval systems have partially solved this problem but prioritization and personalization of information were absent, which has increased the demand for recommender systems that deal with the problem of information overload by filtering vital information fragment according to users preferences and interests [1][2]. Recommender system predicts the future preferences of the user based on the user's profile and thus is proved to improve decision-making process [3]. The need to use efficient recommendation techniques within a system is growing so as to provide relevant and dependable recommendations for users. The use of efficient and accurate recommendation techniques such as content filtering, collaborative filtering, and hybrid approach provides a good and useful recommendation to its individual users.

I.I CONTENT FILTERING

Content-based recommender systems work with profiles of users that are created at the beginning. A profile has information about a user and his taste. Taste is based on how the user rated items. Generally, when creating a profile, recommender systems make a survey, to get initial information about a user in order to avoid the new-user problem. In the recommendation process, the engine compares the items that were already positively rated by the user with the items he didn't rate and looks for similarities. Those items that are similar to the positively rated ones will be recommended to the user. Figure 1 shows an example. We see that there is a movie "DABANG 2" similar to the movie "DABANG" that the user positively rated. The user hasn't rated "DABANG 2" so it will be recommended him/her.

I.II COLLABORATIVE FILTERING

The idea of collaborative filtering is in finding users in a community that share appreciations. If two users have same or almost same rated items in common and then they have similar tastes. Such users build a group or a so-called neighborhood. A user gets recommendations to those items that he/she hasn't rated before, but that was already positively rated by users in his/her neighborhood.

Figure 2 show that all three users rate the movies positively and with similar marks. That means that they have similar taste and build a neighbourhood. The user A hasn't rated the movie "Titanic", which probably means that he hasn't watched it yet. As the movie was positively rated by the other users, he will get this item recommended. As opposed to simpler recommender systems where recommendations based on the most rated item and the most popular item methods, Collaborative recommender systems care about the taste of the user. The taste is considered to be constant or at least change slowly.

Movies	Chennai Express	Golmaal	Dabang	Madras Cafe
Ratings	9	8	10	7

Movies	Genre	Language	Leading Actor	Year
Dabang	Action	Hindi	Salman Khan	2011
Dabang 2	Action	Hindi	Salman Khan	2012
Rowdy Rathore	Action	Hindi	Akshay Kumar	2012
Robin Hood	Drama	English	Russel Crowe	2010
...	...	...	...	...

Figure 1 – Content filtering example

Movie Users	Madras Cafe	Dabang	Dabang 2	Rowdy Rathore	Titanic
User 1	7	10	9	8	...
User 2	5	10	9	7	10
User 3	5	8	8	7	9

Figure 2 – Collaborative filtering example

Collaborative filtering can be of three types: User-based, Item-based and Hybrid-based approaches.

USER-BASED COLLABORATIVE RECOMMENDER SYSTEM: If the certain majority of the customers has the same taste then they join into one group. Recommendations are given to user based on an evaluation of items by other users form the same group, with whom he/she shares common preferences. If the item was positively rated by the community, it will be recommended to the user. Thus in the user-based approach the items that were already rated by the user before play an important role in searching a group that shares appreciations with him.

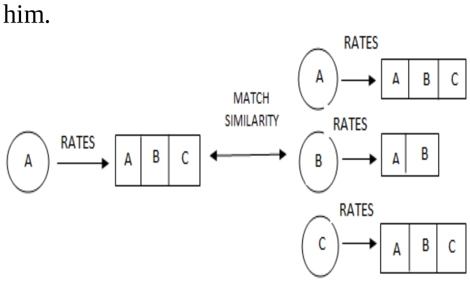

Figure 3 – User-based collaborative filtering example

ITEM-BASED APPROACH: Taste of users remains constant or change very slightly similar items build neighbourhoods based on appreciations of users. Afterward, the system generates recommendations with items in the neighbourhood that a user would prefer.

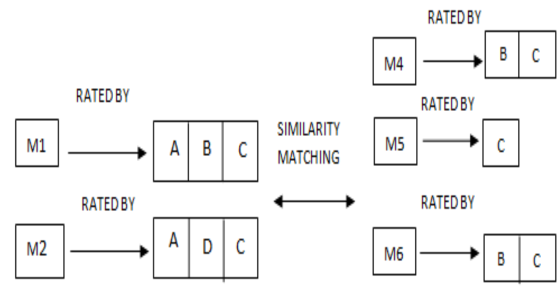

Figure 4 – Item-based collaborative filtering example

HYBRID RECOMMENDATION APPROACH: For better results, some recommender systems combine different techniques of collaborative approaches and content-based approaches. Using hybrid approaches we can avoid some limitations and problems of pure recommender systems, like the cold-start problem. The combination of approaches can proceed in different ways:

1) Separate implementation of algorithms and joining the results.
2) Utilize some rules of content-based filtering in a collaborative approach.
3) Utilize some rules of collaborative filtering in content-based approach.
4) Create a unified recommender system that brings together both approaches.

The organization of paper is: Section II gives related work; Section III presents proposed work, Section IV gives results of the work.

II. RELATED WORK

Table 1 shows the summary of work done in recommendation systems.

Sr. No.	Cita--tion	Category	Method	Research Contribution
1	[4]	Traditional approach	Collaborative filtering	Commonly-used and successfully-deployed recommendation approaches
2	[5]		Neighborhood -based	Focus on finding similar users or items for recommendations.
3	[6]		User-based	Predict the ratings of active users based on the ratings of similar users found

4	[7]		Item-based	Predict the ratings of active users based on the computed information of items similar to those chosen by the active user
5	[6]	Trust-based	Ontology	Runs on a server with the knowledge distributed over the network in the form of ontologies, and employs the Web of trust to generate the recommendations.
6	[8]		Collaborative Filtering	The experiments on a large real dataset show that this work increases the coverage (number of ratings that are predictable) while not reducing the accuracy (the error of predictions).
7	[9]		Standard Collaborative Filtering	The experimental analysis shows that these trust information can help increase recommendation accuracy.
8	[10]		probabilistic graphical model	Experimental analysis shows that this method generates better recommendations than the traditional collaborative Filtering Algorithms.
9	[11]	Social Recommender	Text-based predictor	The experimental results show that incorporating social contextual information can help improve the accuracy of review quality prediction especially when the available training data is sparse.
10	[12]		topic modelling and social network analysis	The proposed method can be applied to a wide range of text mining problems such as author-topic analysis, community discovery, and spatial text mining.

Table 1 - Summary of work done in recommendation systems.

III. PROPOSED WORK

Calculation of similarities is the approach common to all the algorithms for the recommendation. There are basically two distinct sections in any recommendation setup: Similarity calculation and Prediction.

The collaborative filtering uses the above model wherein it calculates the similarities of opinions of users and in turn, generates a recommendation to the user based on results. With the tremendous increase in the amount of data associated with a user (or item), the range of "opinions" mentioned above has increased manifold. Therefore not only the data provided by the user but the data "associated" to the user is also taken into account for calculating similarities using machine learning.
The collaborative filtering technique makes use of Pearson's Coefficient method to calculate the similarity between users as in memory based filtering as given below [13].

$$\mathrm{simil}(x,y) = \frac{\sum\limits_{i \in I_{xy}} (r_{x,i} - \bar{r_x})(r_{y,i} - \bar{r_y})}{\sqrt{\sum\limits_{i \in I_{xy}} (r_{x,i} - \bar{r_x})^2 \sum\limits_{i \in I_{xy}} (r_{y,i} - \bar{r_y})^2}}$$

where x and y are two distinct users, $r_{x,i}$ and $r_{y,i}$ are ratings given by x and y respectively to an item 'i' that belongs in set of all items rated by both the users that yields a similarity value that lies between -1(no similarity) to +1(completely similar).
Another approach is model-based filtering wherein models like Bayesian Networks, Neural networks, and other such approaches are used to predict the items to be recommended to a user. The collaborative filtering can also be item based, unlike above two methods. Here we first find the items similar to those rated by the user and compute the similarity between them, in two distinct steps, similarity and prediction. Implementation is carried out on movie dataset using Python, numpy, scipy, matplotlib, Flask (for web service), and SQLite. Our work is basically a recommendation system which uses user-to-user similarities using Pearson coefficient. Fist we calculate user-user similarity matrix based on their ratings, then for recommending movies to user X, we find k most similar users to X, based on their similarity score. Then we pick top M movies rated by each of this user, find normalized scores for each of those movies multiply with similarity score, then finally pick top n from these and recommend to the user.

IV. RESULTS

This section shows the results of the work implemented using collaborative filtering and reverse collaborative filtering methods.

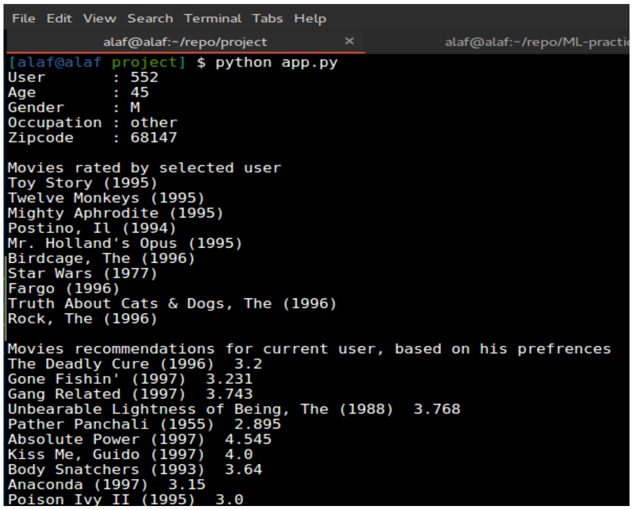

Figure 5 – Collaborative Filtering: Movies rated by selected user and recommendations for the current user based on his preferences

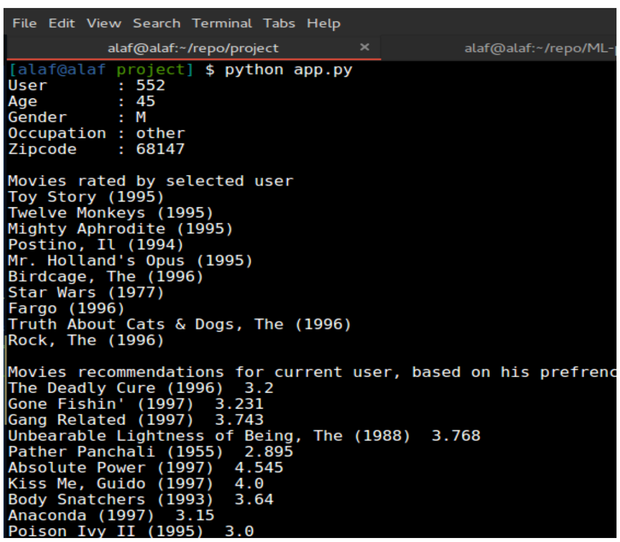

Figure 6 – Reverse Collaborative Filtering: Movies rated by selected user and recommendations for the current user based on his preferences

V. CONCLUSIONS

Recommender systems open new opportunities for retrieving personalized information on the Internet. It also helps to alleviate the problem of information overload which is a very common phenomenon with information retrieval systems and enables users to have access to products and services which are not readily available to users on the system. This paper discusses the efficient recommendation techniques such as content filtering, collaborative filtering and hybrid approach to provide useful recommendation to its individual users. It implements collaborative filtering that predicts the future preferences of the user based on the user's profile. To address the cold start problems in recommendations systems, it implements reverse collaborative filtering.

References

[1] Konstan JA, Riedl J. Recommender systems: from algorithms to the user experience. User Model User-Adapt Interact 2012;22:101–23.

[2] Pan C, Li W. Research paper recommendation with topic analysis. In Computer Design and Applications IEEE 2010;4,pp. V4-264.

[3] Pathak B, Garfinkel R, Gopal R, Venkatesan R, Yin F. Empirical analysis of the impact of recommender systems on sales. JManage, Inform Syst 2010;27(2):159–88.

[4] Z. Huang, H. Chen, and D. Zeng. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. ACM Trans. Inf. Syst., 22(1):116–142, 2004.

[5] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In Proc. of UAI '98, 1998.

[6] R. Jin, J. Y. Chai, and L. Si. An automatic weighting scheme for collaborative filtering. In Proc. of SIGIR '04, pages 337–344, Sheffield, United Kingdom, 2004.

[7] P. Bedi, H. Kaur, and S. Marwaha. Trust-based recommender system for the semantic web. In Proc. Of IJCAI'07, pages 2677–2682, 2007.

[8] P. Massa and P. Avesani. Trust-aware collaborative filtering for recommender systems. In Proceedings of CoopIS/DOA/ODBASE, pages 492–508, 2004.

[9] J. O'Donovan and B. Smyth. Trust in recommender systems. In Proc. of IUI '05, pages 167–174, San Diego, California, USA, 2005.

[10] T. Hofmann. Collaborative filtering via Gaussian probabilistic latent semantic analysis. In Proc. Of SIGIR '03, pages 259–266, Toronto, Canada, 2003.

[11] Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi. Exploiting social context for review quality prediction. In Proc. of WWW '10, pages 691–700, Raleigh, North Carolina, USA, 2010.

[12] Q. Mei, D. Cai, D. Zhang, and C. Zhai. Topic modeling with network regularization. In Proc. Of WWW '08, pages 101–110, Beijing, China, 2008.

[13] Shardanand, U. and Maes, P., 1995, May. Social information filtering: algorithms for automating "word of mouth". In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 210-217). ACM Press/Addison-Wesley Publishing Co.

Authors Profile

Ms. Saniya Zahoor pursed Bachelor of Engineering from University of kashmir in 2011 and Master of Enginering from Pune University in year 2015. She has published 19 research papers in reputed jourals, international conferences and book chapters in IEEE, Springer, Elsievier and ACM. Her main research work focuses on Internet of Things, Personalization, Big Data Analytics, Data Mining.