Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 132 (2018) 1678–1684

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2018)

Generating Top-N Items Recommendation Set Using Collaborative, Content Based Filtering and Rating Variance

Anand Shanker Tewari[a,*], Jyoti Prakash Singh[a], Asim Gopal Barman[b]

[a]*Department of Computer Science and Engineering, National Institute of Technology Patna, Patna, India-800005*
[b]*Department of Mechanical Engineering, National Institute of Technology Patna, Patna, India-800005*

Abstract

The main purpose of any recommendation system is to recommend items of users' interest. Mostly content and collaborative filtering are widely used recommendation systems. Matrix factorization technique is also used by many recommendation systems. All these techniquesproduceconsiderably bigger recommendation list, althoughusers generallyprefer to see fewer recommendations. It means users are interested in smaller recommendations list having items of their interest. To realize this objective, the proposed approach generates smaller top-*n* item recommendations list by placing users' unseen items in recommendation listand thus attaining high precision value. The proposed approach uses content based filtering and collaborative filtering collectively. The proposed recommendation system uniquely finds popularity of all items among users in the form of weights. It also uses the rating variance of different items to generate more effective recommendations.The experimental results shows that proposed recommendation system has better precision, even for smaller number of recommendations when compared with other benchmark recommendation methods.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2018).

* Corresponding author. Tel.:+91-898-705-1537.
E-mail address:anand@nitp.ac.in

Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2018).
10.1016/j.procs.2018.05.139

Keywords: Collaborative filtering;Recommendation system; Rating variance; Content based filtering; E-commerce

1. Introduction

Now with the expansion of the Internet, e-commerce has reached at almost every corner of the world. Almost all e-commerce websites have hundreds of products available online andtheir Recommendation System (RS). It is the responsibility of the RS to recommend right products to the user. These RS not only helps in retaining existing customers but also increases the sales profit by generating meaningful recommendations to users. The two popular recommendation techniques are content based filtering and collaborative filtering respectively. Content Based Filtering (CBF) uses products' descriptions of used products to find other products that may seem interesting for the target user [1]. Content based recommendation methods matches users profiles with items profiles and recommends only those items that have high similarity with users' interest [2]. However CBFisunable to do quality assessment and does not generate serendipitous recommendations to the target user [3].

Collaborative Filtering (CF) overcomes the limitations of the content based filtering. CF finds interesting items for the target user, on the basis of ratings of other likeminded users in the same domain. CF systems are usually based on human judgmentsusually given as ratings. CF does not need items' descriptionsto generate recommendation for the target user [4].

Matrix Factorization (MF) is another popular recommendation technique. It is based on the singular value decomposition of the rating matrix. It generates dense approximate rating matrix from the original rating data[5].

Generallybuyers like small number of recommendations frome-commerce websites. Most recommendation systems generates big recommendation list that may have users' preferred items at the bottom. The proposed RS generates smaller recommendation list with high precision value. It means more interesting itemsare available in relatively smaller recommendation list. The proposed approach uses content based filtering to create dynamic profile of users and collaborative filtering as a *classifier*. The proposed RS calculates the popularity of different itemsamong other users, in the form of *weights* and which helps in generating high precision recommendations [6].It also finds the rating variance of all items. All these techniques collectively help in placing those items in smaller recommendation set that may be preferred by the target user. The complete structure of the paper is organized as follows: Section 2 discusses background knowledge, Section 3 elaborates proposed approach, Section 4 shows practical implementation and evaluation of the proposed work and finally Section 5 concludes the paper.

2. Background knowledge

2.1. Content based filtering

CBFis the natural process running all the times in human brain. It is not limited to electronic documents [7]. Example: Whenever people are reading details printed on the packaging of product, they usually read only selective portions of packaging based on their interest rather than reading the complete details.Recommendation systems based onCBFconstructs the profile of each user by analyzing the content of items preferred by him the past and matches it with other items. Then, CBF recommends only those items that have high degree of similarity to items that user has likedearlier [1, 8]. CBF approach to recommendation has come from information retrieval and information filtering research [9, 10]. Suppose $D = (doc_1, doc_2, ... doc_N)$ is a collection of documents and $T = (t_1, t_2, ..., t_n)$ be the set of keywords in the collection. Each document can be represented by a vector in n-dimensional vector space. Then, the cosine based similarity between the two documents doc_1 and $doc_2$is calculated using Eq. (1).

$$\text{Similarity}(\overrightarrow{doc_1}, \overrightarrow{doc_2}) = \text{cosine}\,(\overrightarrow{\mathrm{doc}_1}, \overrightarrow{\mathrm{doc}_2})$$

$$\text{cosine}(\overrightarrow{doc_1}, \overrightarrow{doc_2}) = \frac{\overrightarrow{doc_1}.\overrightarrow{doc_2}}{|\overrightarrow{doc_1}|\times|\overrightarrow{doc_2}|} \quad (1)$$

2.2. Collaborative filtering

Recommendations based on content based filtering has certain shortcomings, like it cannot differentiate between good and bad items if both have same set of attributes. CBF does not generate serendipitous recommendations to users [3]. Collaborative filtering judges items' qualityusing ratings [11]. CF creates the profile of each user by collecting his items'ratings from website and compares it with other users' ratings to find users with similar interests. Then the ratings of these similar users are used to generate recommendations for the target user. Similarity ($Sim_{m,n}$) between the two users$m$ and n, can be calculated using Pearson's correlation formula as shownin Eq. (2).

$$Sim\ (m,n) = \frac{\sum_{a \in Imn}(Rm,a - Am)(Rn,a - An)}{\sqrt{\sum_{a \in Imn}(Rm,a - Am)^2 \sum_{a \in Imn}(Rn,a - An)^2}} \quad (2)$$

Where $R_{m,a}$ is item $a$rating by user m, similarly $R_{n,a}$ is item $a$rating by user n, A_m and A_n are average ratings of user m and n respectively and $I_{m,n}$ is commonly rated items by both users.Thepredicted rating(P_{mk}) of the target user m to the target item k is calculated as shown in Eq. (3).

$$P_{mk} = A_m + \frac{\sum_{n=1}^{c}(Rn,k - An)^* Sim(m,n)}{\sum_{n=1}^{c} Sim(m,n)} \quad (3)$$

Where $R_{n,k}$ is item $k$rating by similar user n, c is the total number of similar users to the target user$m$ and$Sim(m,n)$ is the similarity between user m and user n.

3. Proposed Approach

The proposed RS usesits unique five basic building blocks to generate recommendations for the target user. These blocks are Profile Builder, Similarity Finder, Collaborative Classifier, Item Weight and Variance calculator and Final Recommender. The functioning of each block is described as follows:

3.1. Profile Builder(PB)

The PB block constructs the profile of every user and item. It creates the item profile by set of keywords, usually provided by sellers in e-commerce website, for describing an item. The PB block also creates the profile each user in the form Keywords Vector (KV). User's keywords vector is constructed from items profiles of those items that are used by the target user. Each user's KV stores information regarding user's preferred past and present items[12]. The PB block dynamically updates every user's KV after regular intervals.

3.2. Similarity Finder (SF)

The SF block collects users' keywords vectors from the PB block. Afterwards the SF block finds the similarity of every user with other user using Eq. (1), this equation can be rewritten as shown in Eq. (4), where UKV_1 and UKV_2 are keywords vectors of $user_1$ and $user_2$ respectively.

$$\text{cosine}(\overrightarrow{UKV_1}, \overrightarrow{UKV_2}) = \frac{\overrightarrow{UKV_1} . \overrightarrow{UKV_2}}{|\overrightarrow{UKV_1}| \times |\overrightarrow{UKV_2}|} \quad (4)$$

3.3. Collaborative Classifier (CC)

Collaborative classifier block uses CF to generate recommendations. This block works as a classifier. The CC block selects k most similar users to the target user with the help of SF block for target user's unseen items ratings prediction. Then CC block predicts the ratings of items not used by the target user using Eq. (3). Next itclassifies all predicted items for the target user into two categories *like* and *dislike*[13]. The CC block arranges all existingitems ratings of the target user into four quartiles. Next CC block keeps allthose predicted items in *like*category where their predicted ratings are greater than or equal to fourth quartile ratings of the target user.

Table 1. Items weights calculation table.

Item_Id (i)	Average Item Rating (AIR)	Number of ratings received by an item
I_2	3.1	21
I_4	5	17
I_5	1.9	10
I_7	4	16

3.4. Item Weight and Variance calculator(IWV)

The IWV block helps in finding the popularity of different items among all users in the form of *weights*. It also calculates the rating variance of different items across different users [14].The item weight of any item is calculated using Eq. (5), where AIR_i is the average item rating of an item i, $Rating_{max}$ is the largest rating in the given rating scale, $Count_i$ is the number of users who have rated an $item_i$ and $Count_{max}$ is the highest value of any $Count_i$ for any item. Example: Table 2 shows an instance of average item ratings of different items and number of users who have given ratings to particular item i. For this instance of data *item weight* of an item I_7 is calculated by putting the values in Eq. (5). The values of AIR_7 is 4, $Count_7$ is 16, $Rating_{max}$ is 5 and $Count_{max}$ is 21.

$$IW_i = \frac{AIR_i}{Rating_{max}} \times \frac{Count_i}{Count_{max}} \quad (5)$$

The item ratings*variance*is calculated using standard deviation as shown in Eq. (6), where SD_i is rating standard deviation of an $item_i$.

$$SD_i = \sqrt{\frac{\sum_{k=1}^{Count_i} (x_k - AIR_i)^2}{Count_i}} \quad (6)$$

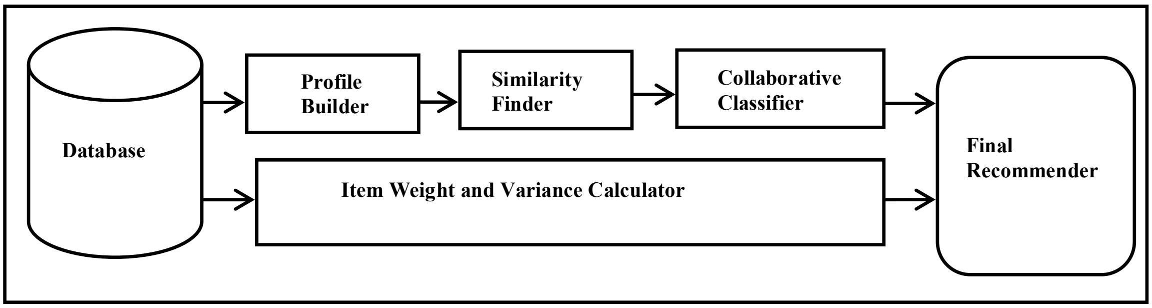

Fig. 1. Block diagram of the proposed approach.

3.5. Final Recommender(FR)

Final recommendations for the target user are generated by the FR block. It collects inputs from the CC and IWV blocks. The CC block generates binary output. All items that are in *like* category in the context of a target user, CC block generates *1* as output and for other items it generates *0* as output. The FR block collects *items weights* and their respective *variance* from the IWV block. FR block finally generates Final Recommendation Value (FRV) for each item as shown in Eq. (7). Afterwards FR block generates item recommendation list for the target user in decreasing order ofitems' FRV values. The block diagram of the proposed approach is shown in Fig. 1.

$$FRV_i = (CC_i - SD_i) + IW_i \quad (7)$$

4. Practical Implementation and Evaluation

The proposed RS is build using Java7. Users' profiles, ratings and other useful data are stored in the MySql database. The proposed RS is tested on live dataset of the website myopinions.in, having 4100 ratings given by 640 genuine users about 1200 movies and books.

The primary concern of each e-commerce website is selling, so the proposedRS is tested on *precision* decision support accuracy metric. To evaluate the performance of the proposed RS it is compared against state-of-the-art recommendation methods like CBF, CF and MF technique. The *precision* metric,evaluate the recommendation system, helpfulness in selecting better items among all available items [15]. Precision is the fraction of correct number of recommendations to the total number of recommendations generated, it is defined as follows:

$$Precision = \frac{Number\ of\ Relevant\ recommendations\ generated}{Total\ number\ of\ recommendations\ generated} \quad (8)$$

To select the value of top-n,a survey has been conducted among 300 people and found that most people prefer to see 10-25 recommendations. Users' avoid bigger recommendation list and thus performance of the proposed RS istestedfor top-25 items(n=25). Firstly user based collaborative recommendations are generated for the target user with the help of CC block. Personalized recommendations are generated by CC block for the target user with the help of similar users. The *precision* comparison graph of top-25 user based collaborative recommendationsby varyingnumber of neighbours (k) is shown in Fig. 2.

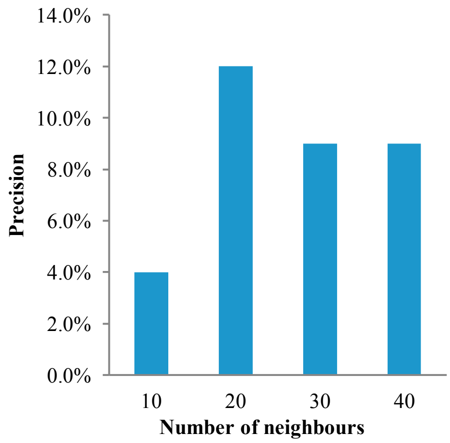

Fig. 2. Precision comparison graph of collaborative recommendations with different number of neighbours.

It is clear from the graph shown in Fig. 2. that recommendations generated by 20 nearest neighbours have higher precision value on the used data, so the value of $k$is selected as 20 for the CC block to generate further recommendations for the target user. Now IWV block has generated the *item weights* and *variance* of all items. Afterwards FR block has generated the final recommendations for users. The *precision* comparison graph of proposed approach with other standard recommendation approaches is shown in Fig. 3.

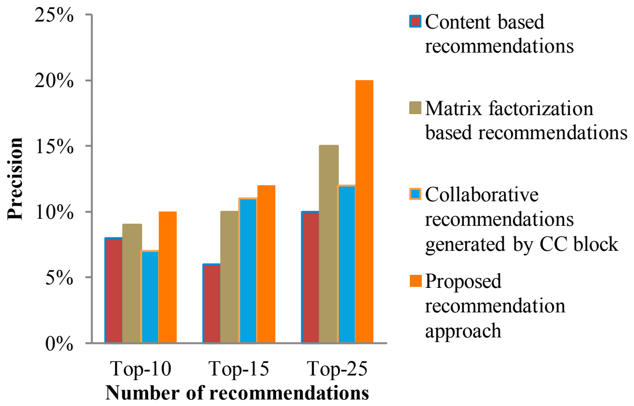

Fig. 3. Precision comparison graph of different recommendation techniques with the proposed approach.

5. Conclusion

Almost all e-commerce RS generates big top-*n* items recommendation list that may have interesting items at the end of the recommendations list. In general users' interest starts decaying with each successive item in the recommendation list.So there are high chances that user may not look at some interesting items at the bottom of the bigger list. The proposed approach with the help of its five unique building blocks generates smaller recommendations list in such a manner that more interesting items for the target user comes at the beginning of the list.The result shows that proposed work gives higher precision than other benchmark recommendation methods even for smaller recommendation list. This approach needs only items descriptions and users rating data for generating effective recommendations, so the proposed RS can be used by any e-commerce websites. In future it can be extended by using data related to target user's friends, tags, etc., generated from online social networks and other platforms.

References

[1] Pazzani M.J., and D. Billsus. (2007) "Content-Based Recommendation Systems." in Springer Lecture Notes in Computer Science *The Adaptive Web***4321**: 325-341.

[2] Adomavicius G., and A. Tuzhilin. (2005) "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions." *IEEE Transactions on Knowledge and Data Engineering***17(6)**: 734–749.

[3] Shardanand, U., and P. Maes. (1995) "Social Information Filtering: Algorithms for Automating 'Word of Mouth'.", in proceedings of the *SIGCHI Conf. on Human Factors in Computing Systems*: 210-217.

[4] Kourosh, Modarresi (2016) "Recommendation System Based on Complete Personalization." *Procedia Computer Science***80**: 2190-2204.

[5] Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. (2000) "Application of dimensionality reduction in recommender systems-a casestudy." in *ACM WebKDD Workshop*.

[6] Tewari, A.S., and A.G. Barman. (2018) "Sequencing of items in personalized recommendations using multiple recommendation techniques." *An International Journal: Expert Systems with Applications***97**: 70-82.

[7] Foltz, P. W., and S. T. Dumais. (1992) "Personalized information delivery: an analysis of information filtering methods",*Communications of the ACM***35(12)**: 51-60.

[8] Claypool, M., A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. (1999) "Combining Content-Based and CollaborativeFilters in an Online Newspaper."in proceedings of*ACM Workshop Recommender Systems: Algorithms and Evaluation.*

[9]Baeza-Yates, R., and B. Ribeiro-Neto. (1999) "Modern Information Retrieval" in Addison-Wesley.

[10] Belkin, N., and B. Croft. (1992) "Information Filtering and Information Retrieval:two sides of the same coin?"*Communications of the ACM* **35(12)**:29-38.

[11] Herlocker, J. L., J. A. Konstan, A. Borchers, and J. Riedl. (1999) "An Algorithmic Framework for Performing CollaborativeFiltering." in Proc. of 22nd *International ACM SIGIR Conference on Research and Development in Information Retrieval*: 230-237.

[12]Tewari, A.S., and A.G. Barman. (2017) "Collaborative Recommendation System Using Dynamic Content based Filtering,Association Rule Mining and Opinion Mining."*International Journal of Intelligent Engineering and Systems* **10(5)**: 57-66.

[13] Basu, Chumki, H. Hirsh, and W. Cohen. (1998) "Recommendation as Classification: Using Social and Content-Based Informationin Recommendation." in 15th*National Conf. on Artificial Intelligence*: 714-720.

[14]Kwon, YoungOk (2008) "Improving top-n recommendation techniques using rating variance." in proceedings of *ACM Conference on Recommender Systems*: 307-310.

[15]Isinkaye, F.O., Y.O. Folajimi, and B.A. Ojokoh. (2015) "Recommendation systems: Principles, methods and evaluation."*Egyptian Informatics Journal***16 (3)**: 261-273.