Generating Quality Items Recommendation by Fusing Content based and Collaborative filtering

Anand Shanker Tewari, Aleshea S.J.

Abstract: Recommendation system has become an inevitable part of our life. It has already spread its prominence in various fields like movies, music, news, article recommendations etc. Due to the influence of social media, data is streaming from all over the Internet. Collect the relevant information from chunks of data available has become much difficult. Recommender systems guides in filtering data to get the relevant information. Commonly used recommendation approaches are content based filtering and collaborative filtering. Each approach has its own limitations. The hybrid approach combines the advantages of both the approaches. In this paper, we have tried to enhance the quality of the items recommendation system by fusing both content based and collaborative filtering uniquely. The experimental results are compared with that of other traditional approach using precision and recall evaluation measure. The comparison results show that our approach has 10% better precision for top-10 recommendations than other established recommendation technique.

Index Terms: E-commerce, Content based filtering, Collaborative filtering, Hybrid Recommendation System.

I. INTRODUCTION

Day by day the Internet is getting overloaded with a large amount of data. This makes difficult for users to find out what they exactly need. The relevant ones get meld with that of irrelevant. Recommendation system helps to refine data. Nowadays a lot of researches are going on in the field of recommendation system on how to refine data. There are many approaches to recommend items. Content based and collaborative filtering are the most popular approaches used. Content based filtering technique recommends items that are similar in content to the users’ and items’ profile construction. Construction of users’ profile is done with the help of the content of items that were previously rated [1]. For example to suggest the user with items using this approach, first fetches information from the user's profile. According to the profile’s information recommendation of items are generated. Constructing a profile to a new user is difficult. Recommending items to a completely new user is thus a difficult task [2].

Collaborative filtering is another popular approach where the recommendations are generated based on the ratings. There are two approaches for collaborative filtering. Like-minded user tends to prefer the same items. The users of the same tastes are grouped together based on the rated items. Items that have vacant ratings are recommended from the profile of similar users’ to the target user. Another approach of collaborative filtering is based on similar items [3]. It recommends items that are similar to the items, the user have rated previously [4].

Hybrid approach combines both the content based filtering and collaborative filtering approach. Combining the advantages of both the approaches makes the recommendation better [5].

The proposed approach is based on the hybrid approach. It has combined the content and collaborative filtering approach. The proposed approach has specialized weights calculation blocks that helps in placing relevant items at top-n positions. Thus this approach gives better precision and recall. The following sections are structured as: Section 2 addresses the related works followed by the section 3 briefly explains the proposed approach. Section 4 states practical implementation and evaluation and finally section 5 concludes the paper.

II. RELATED WORKS

The collaborative filtering and content based filtering are few prominent techniques. They are widely used in many commercial applications [6]. In [7] paper authors have used collaborative filtering on items’ categories and finds items interestingness, later they combined both these factors to generate recommendations. The Qin, Cao, and Peng have given idea of collaborative filtering recommendations based on weighted items’ categories. The approach has tried to handle sparsity problem to improve the recommendations [8]. Mu, Xiao, Tang, Luo, and Yina have introduced new similarity measure, which strongly depends on the distance function [9]. The Gandhi’s paper has combined the collaborative filtering with the association rule mining to generate recommendations to improve the quality of the recommendation system [10]. Similarly Tewari and Priyanka have also developed a recommendation system for recommending books using user based collaborative filtering and association rule mining [11]. Kardan and Ebrahimi have proposed a novel approach in hybrid recommendation systems [12]. They have applied association rules mining to trace the best similar users. The Puntheanaruk and Chaiwitooanukool have used items attributes similarities and later use these similarities to adjust the values of the predicted ratings [13]. Statistical Relational Learning can be used for a hybrid approach by using the probabilistic dependencies among the attributes [14]. A news utility model was developed to address the problem of considering clicks as the most effective indicator of real user interests by most of all the existing recommendation system.
Social media helps to learn users’ interest faster. This technique is used to recommend occupation to the students using their profile, activity and hobbies [16]. A hybrid approach using data mining techniques for tourism recommendation was proposed by Lucas and Joel [17]. They used associative classification which combines classification and association. Movie recommendation by Wei and Shouxian uses tags and ratings [18] for building a social movie network and a preference topic model. The tags are then extracted, normalized to catch up the user preferences.

### III. PROPOSED APPROACH

The proposed approach consists of mainly six blocks, each block possessing a unique functionality. The first block called User Profile Construction Block, the second one is Similar User Profile Block, third is Item Rating Prediction Block fourth is Preference Weight Calculation. Fifth is the Common Weight Block. The last block is called Final Prediction block. The detailed description of each block is as follows:

1. **User Profile Construction (UPC) Block**

The UPC block constructs the user’s profile of each user based on the ratings of all the items rated by them. The block also fetches all the features of items, rated by the target user and uses these features to constructs the user vector. This vector is filled with ratio of count of particular feature to the sum of frequencies of all the features used by that user. By this way, the block creates the profile for each user to trace out the user’s personal preferences. An example for the users’ vector table is shown in table 1.

<table>
<thead>
<tr>
<th>Users</th>
<th>Comedy</th>
<th>Romance</th>
<th>.</th>
<th>.</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>U_1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>U_2</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>U_m</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1: Users profile vector table.

2. **Similar User Profile (SUP) Block**

This block finds the users that have same taste with that of target user. The similar users’ profiles are found out using the Pearson correlation similarity \( sim_{u,u} \) as shown in (1).

\[
sim_{u,u} = \frac{\text{cov}(x_u, x_a)}{\text{std}(x_u) \times \text{std}(x_a)}
\]  

(1)

Where \( sim_{u,u} \) is the similarity between two users i.e. user \( u \) and user \( a \), based on ratings given by both users on all common \( n \) items. The similarities between users that maintain a value greater than or equal to certain threshold, \( sim \), are considered as similar users to the target user. The SUP block then places items which all are not rated by the target user but used by his similar users in an interested_items list for the target user.

3. **Item Rating Prediction (IRP) Block**

This block calculates the rating of all items available in the interested_items list of the target user. This is to assign priority to the items that have higher rating and are from more similar users’ to target user. The prediction calculation is done using (2).

\[
R_{u,i} = \bar{r}_u + \sum_{a=1}^{n} \frac{\text{sup}(\text{rating}_{u,a}) \times sim_{u,a}}{\text{sum}_{u,a}}
\]

(2)

Where \( R_{u,i} \) defines predicted rating of item \( i \) for the user \( u \). \( sim_{u,a} \) is calculated from the SUP block. The \( \bar{r}_u \) is the mean rating of the target user \( u \) and the number users similar to the target user is \( n \).

<table>
<thead>
<tr>
<th>Movies</th>
<th>Comedy</th>
<th>Romance</th>
<th>.</th>
<th>.</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>M_1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>M_2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>M_n</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Movie profiles vector table.

Once the ratings are calculated, the IRP block then converts all generated items rating in to two values i.e. 0 and 1 [19]. It performs this operation using below steps:

1.) The block arranges all items ratings of each user in the ascending order.

2.) The block then finds all the ratings given by target user which are in the third quartile and above. The starting point of the third quartile defines the rating_separator or_value.

3.) The number of predicted items’ rating for the target user that are greater than equal to rating_separator_value is updated with 1 and rests of the predicted items’ rating are set as 0.
4. Preference Weight Calculation (PWC) Block

This block computes the weight of all items available in the target users’ interested_items list. The items that have more features as in the target user profile, gets higher preference. Example: Suppose, a movie $i$ that is in the interested_items list of the target user $u$, have genres like action, thriller, adventurous and romance but the target user have genres like action, thriller and adventurous in their profile. So, there are three genres in common. The percentage of each genre from the target user profile vector, that has entry 1 in movie profile vector as shown in table 2, is then added together to find the weight ($w_{ui}$) of that item as given in (3).

$$w_{ui} = \sum_{j=1}^{r} f_{u,j} \times b_{i,j}$$

(3)

Where $f_{u,j}$ is the $j^{th}$ feature value from the user’s vector $u$ that matches with the item’s vector $i$ and $b_{i,j}$ is the binary value of the $j^{th}$ feature of an item’s vector $i$. The $r$ is the number of common features available in both user’s and item’s profile vectors.

5. Common Weight (CW) Block

This block finds the preferences of different items among all other users. This helps the proposed recommendation system to place better items for the target user. The CW block helps in rearranging relevant items in the interested_items list of the target user. The common weight (CW$_i$) of an item $i$ is calculated by using (4).

$$CW_i = \frac{\text{avg}_i \times C_i}{R_{max}}$$

(4)

Where $\text{avg}_i$ is the items $i$ average rating, $R_{max}$ is the maximum rating for the rating scale given. $C_i$ is the total users who gave rating to item $i$ and $C_{max}$ is the maximum $C_i$ value among all items [20].

6. Final Prediction Block

This block takes the inputs from the PWC block, IRP block and CW block. The outputs from all of these blocks together produce a final recommendation for the target user as shown in (5).

$$F\text{Rec}_{ui} = (w_{ui} \times \text{IRP}) + CW_i$$

(5)

Where $F\text{Rec}_{ui}$ is the final recommendation value of an item $i$ for user $u$ and $w_{ui}$ is the preference weight of user $u$ for an item $i$ calculated by the PWC block. The IRP is the output from the item rating prediction block and $CW_i$ is the common weight of an item $i$. The items’ $F\text{Rec}_{ui}$ values are arranged in decreasing order and are recommended to the target user. Block diagram of the proposed approach is shown in Fig. 1.

IV. PRACTICAL IMPLEMENTATION AND EVALUATION

We have conducted several experiments to compare the recommendation quality of proposed approach using live dataset.

A. Experimental Setup

The implementation of the proposed system was done in Java 8. Intel Core i7 machine with 7.6 GB RAM is used for experiments. We have used live dataset of http://www.myopinions.in. It has Indian as well as Hollywood movies. The dataset is having 520 realistic users, 1038 movies and 8227 ratings of those movies. Movies rating scale ranges from 1 to 5.

B. Experimental Evaluation Methodologies

We compared our proposed approach with Precision and Recall evaluation measures. These measures are used extensively in the recommendation systems. The qualities of the results are evaluated by these ratios. The ratio between the total relevant items to the total items recommended is the Precision as shown in equation (6) whereas the ratio of the total relevant items that was recommended to the total relevant items in the items library is given by Recall as shown in (7).

$$\text{Precision} = \frac{\text{relevant items}}{\text{total recommended items}}$$

(6)

$$\text{Recall} = \frac{\text{relevant items}}{\text{total relevant items}}$$

(7)

Fig. 1. Block diagram of the proposed approach.

Fig. 2. Precision comparison of User based collaborative filtering with the proposed approach.
Generating Quality Items Recommendation by Fusing Content based and Collaborative filtering

C. Practical Evaluation

The experiment is started by implementing the UPC block and then Similar User Profile block. The SUP block finds similar users using (1). We calculated the users’ similarity for different threshold i.e. sim, values are as 0.5, 0.6, 0.7, 0.8, and 0.9. The sim value at 0.7 has given better precision and recall than rest of the other values. So we have chosen sim value as 0.7 and created the interested_items list for users.

To predict the rating of all the movies, in interested_items lists are then fed to Item Rating Prediction block for calculation. In the next step the IRP Block finds the rating_separator_value of each user. The predicted ratings of movies are then updated with 1 and 0s.

Common Weight block calculates the popularity of all the movies among users using (4). Preference Weight Calculation block finds the weights of different items according to the users’ and movies’ profile. If there is any match of genre in the movie and user profile then weight is calculated using (3).

The Final Prediction block then finally generates FRecV values for every item in the interested_items list of the target user. These items are recommended in descending order of their FRecV values. The precision and recall value of the generated recommendations are then compared with the established user based collaborative filtering technique. The proposed approach performs better every time than the collaborative filtering approach as visible in fig 2 and 3. The approach has shown a significant precision improvement of 10% for top-10 recommendations and 3% for top-25 recommendations.

V. CONCLUSION

The proposed approach generates better recommendations than traditional recommendation techniques using its unique building blocks. The weight that is introduced in Preference Weight Calculation block enhances to popes the relevant items which were at the bottom of the RS to be in top n number. The approach has tried to improve the precision for smaller recommendation list. We compared our approach with that of user based collaborative filtering. The results proves that the proposed approach outperforms that of user based filtering and help users to fetch the relevant items without much searching throughout the list. This approach requires only users rating and items descriptions for recommendation generation. Thus the proposed RS can be implemented by any movie recommendation system. The work can be improvised by including more data such as tags, demographic information, etc.

ACKNOWLEDGMENT

First of all, let me thank the Almighty for his blessing for the completion of my work. Secondly my guide Dr. Anand Shanker Tewari for his support and encouragement without which the work will not reach to its fulfillment.

REFERENCES


AUTHORS PROFILE

**Anand Shanker Tewari** has done his B. E. and M.Tech in Information Technology. He has done his Ph.D. in Computer Science and Engineering. Currently he is doing research in recommendation system and published many papers.

**Aleesha S. J** has done B.Tech in Information Technology. She is currently doing M.Tech in Information Technology from NIT Patna.