Design & Implementation of Customer Targeted E-Commerce using Recommendation Systems

Keshav Maheshwari

Computer Science, Birla Institute of Technology and Science, Pilani, Dubai Campus

Abstract- The Thesis is concerned with the basic study of E-commerce with recommendation systems enabled**.** Understanding your target demographic is a cornerstone of good business performance management, or BPM**.** Recommendation systems were evolved as intelligent algorithms, which can generate results in the form of recommendations to users. Unfortunately, the traditional collaborative filtering systems cannot make accurate recommendation for the two cases because the predicted item for active user is not consist with the common interests of his providing users, and hence the idea of Hybrid Recommendation System is proposed. There is also an introduction to the implementation of the technology for augmented reality in E-commerce but this area is still researched upon.

Index Terms- E-commerce, Recommendation System, Interest, Preference, Augmented Reality

I. Introduction

Recommendation system are superficially classified into 3 categories This paper discusses the approach of collaborative filtering. Collaborative filtering approach suggests things or items having a similarity coefficient between them. The things that are suggested are measured by those who have a similar interest.

Related work: Collaborative filtering (CF) is a recommending phenomenon used by recommender systems. Collaborative filtering has two aspects, a specific one and a more generic one. In the specific sense, it is defined as the process to make automatic filtering about the likes and dislikes of a user by collecting the information or collection of information by many users. [1]. The unsaid assumption of collaborative filtering is that if user X has same interest as a user Y on a product, then X is more probable to have the same opinion as of Y on a different issue than that of any different user. For Example, A collaborative filtering recommendation system for Web series taste can make predictions about which web series a user shall like, given a half list of that user's interest. Important to know that all these predictions are pointed to user, but the information used is collected from many users. This is different from a simpler approach which is giving a mean point to each product of interest. [3]

Contributions: This paper has three primary research contributions:

1. Identification and analysis on limitation of collaborative filtering based on user for Multiple-interests and Multiple-content recommendation.
2. Presentation of a hybrid collaborative filtering method, collaborative filtering based on item and user algorithm, to improve recommendation accuracy for Multiple- interests and Multiple-content recommendation.

II. Problem Analysis

Shortcomings of Collaborative Filtering: It fails to provide personalization in recommending products to the user. Even more, CFU is better at providing recommendation, as the results of some experiments are show. Another reason which makes CFI better than CFU is that if users have different opinions, likes and dislikes or items differ in features and attributes defined by us as "Multiple-Interests" & "Multiple-Content" issue.

Problem: As per CFU, something called as “user preference” is used to know the prediction value for item which is being targeted. There is a difference between rated items and predicted items and that is that, predicted items are on different content and rated items are on same content, but one user can have different types of interest or we call “Multiple Interest” and items have different types of content which we call “Multiple Content”. The predicted item has no connection whatsoever with the rated items and that is a truth which makes it believable. Consider, there is a chance that a user is inclined towards “Cricket” and “German” still provide ratings only to Cricket items. Hence, the “interest preference of Cricket items user is taken in consideration for prediction of German items, which will still cast a doubt on precision. [2].

Consider the given table, it’s a data matrix for user and item, comprised of 7 users and 6 items. We can see that the 3 items which is Y1(GERMAN), Y3(GERMAN) and Y6 (GERMAN) are same when we talk about content but different when we talk about items. Similarly, 3 items which is Y2(CRICKET), Y4(CRICKET) and Y5 (CRICKET) are same on the content cricket but entirely different from content Y1, Y3 AND Y6.

Now, consider we want to predict rating (r_{76}) of Row 7 Col. 6 i.e. user X7 for item Y6. {The user X7 = “active” and item Y6 = “predicted”}. I assume every user is surrounded by 3 neighbors.

If we use CF-u to make the prediction for this problem, we can easily say that users X4,X5 and X6 are nearest neighbours for X7 as their rating style is same to that of X7.It is simple to calculate the value, r_{76}=2. Focussing on the reason that why X4,X5 and X6 are the neighbours of X7 is because their interest is same which is cricket.

Both Cricket & German are not related & dependent but still, we are using the perference of users with items having relation with Cricket (Y2, Y4, Y5) to predict “interest preference” on item having relation with GERMAN (Y6), therefore this creates a question on the prediction that we made and doubts for accuracy.

	ITEM					
USER	Y1(GERMAN)	Y2(CRICKET)	Y3(GERMAN)	Y4(CRICKET)	Y5(CRICKET)	Y6(GERMAN)
X1	4	2	3	4	6	6
X2	4	2	3	4	6	6
X3	4	2	3	4	6	6
X4	2	6	4	4	2	2
X5	3	6	3	4	3	2
X6	4	6	2	4	3	2
X7	4	6	3	5	3	?

Table 1: “An example of user/item data matrix”

To elaborate the perspective, consider an advanced situation. For the same table, assume that no value is assigned to X4, X5 and X6 for the items Y1 and Y3, we will get a new table, the question remains the same, which is r_{76}=?

CF-u will work normally, Neighbors of X7 will be X5, X6 and X4. Preference information is used to identify the value of rating for X7 for Y6 as X5, X4 and X6 show the same rating as active use X7. But there is difference in content for the predicted item Y6 (German) than items Y2, Y4,Y5 (cricket) for users X4,X5,X6,X7. Hence, the prediction has some flaws.

	ITEM					
USER	Y1(GERMAN)	Y2(CRICKET)	Y3(GERMAN)	Y4(CRICKET)	Y5(CRICKET)	Y6(GERMAN)
X1	4	2	3	4	6	6
X2	4	2	3	4	6	6
X3	4	2	3	4	6	6
X4	...	6	...	4	2	2
X5	...	6	...	4	3	2
X6	...	6	...	4	3	2
X7	4	6	3	5	3	?

Table 2: “Missing Value Table”

Solution: For the discussed model problem, If interest preference for every users for items GERMAN is used into determine the neighbours of the active user X7 (“meaning that, similarity of user X7 to the rest of these users are calculated accordingly by considering the ratings of rest users for Y1 & Y3, for German), we conclude that X1, X2. X3 are the neighbours of active user U7. This makes the predicted value, p_{76}=5 as “interest preference” of active user X7 & its neighbours X1, X2. X3 for German is similar, & rating of neighbours X1, X2. X3 for item Y6 can be utilized to predict rating of active user U7 for item Y6.

According to above analysis, we can get a conclusion that active user should has common interest in predicted item with its neighbour users. So, computation of similarity of active user to other users should be based on items related to predicted item, not on all items, more being not based on items not related to predicted item. It means that, for different predicted item, neighbours of the same active user are different. It can guarantee that predicted item and items used to predict for predicted item are similar on content, which can lead to improvement for CF-U. According the solution, we present a hybrid collaborative filtering method, collaborative filtering based on item and user algorithm,

III. Hybrid collaborative filtering method

According to above analysis, we present a hybrid collaborative filtering method, collaborative filtering based on item and user (CF-IU), by combining collaborative filtering based on item (CF-I) and collaborative filtering based on user (CF-U). The new method is adaptive to Multiple-interest and Multiple-content recommendation.

Algorithm:

Some denotions,

- $r_{x,y}$—x’s rating for y with interest preference;
- 2) “$I_{r,x}$={x|$R_{x,y}$ is not null, y∈1,2,.,n}”,x rated itemsets;
- “$I_{P,y}$={x|$R_{x,y}$ is null, y∈1,2,.,n}”, non x rated itemsets;

❖ Inputs:

- Rating of m users for n items (allowing null rating);
- Threshold s,w and r for item, user and rating respectively
- simITEMNum—Max no. of same items for predicted item;
- neiNum—the max no. of neighbours for active user;

User	Item					
	I1	**I2**	…	**Ij**	…	**In**
U1	R_{11}	R_{12}	…	R_{1j}	…	R_{1n}
U2	R_{21}	R_{22}	…	R_{2j}	…	R_{2n}
…	…	…	…	…	…	…
Ui	R_{i1}	R_{i2}	…	$\mathbf{R_{ij}=?}$	…	R_{in}
…	…	…	…	…	…	…
Um	R_{m1}	R_{m2}	…	R_{mj}	…	R_{mn}

❖ Outputs:

The L items that are interesting in the item set $I_{p,a}$ for user a is made as a recommendation to active user

- Computing procedure:

1. given the target item $y \in I$, the similarity calculations item *y*& other items $z(z \in 1,2,., y-1, y+1, n)$—Sim(y,z);Similarity may be measured by Pearson correlation coefficient, cosine & so on;
2. According to Sim(y,z) $(z \in 1,2,., y-1, y+1, n)$, determine similar item sets to target item y–SI_y;

   Two methods are used to determine SI_y:

   - Thresholding: by setting an outright item closeness edge s, to decide comparable things SI_y to target thing with supreme comparability Sim(y,z) more prominent than s.
   - Best-things: getting the max simITEMNum closeness from all Sim(y,z), $(k \in 1,2,., y-1, y+1, n)$, and relating simITEMNum items structure SI_y;

3. Compute similarity $w_y(a,i)$ of active user a to other user $x(x \neq a)$ based on item sets SI_y;
   1. identify mutual rated item sets $CSI_y(a,x)$ of

      user a & user x;

   2. Compute similarity $w_j(a,x)$ of active user a to other user $x(x \neq a)$ based on CSIy(a,x) using "pearson correlation".

4. According to similarity $w_y(a,x)$, determine neighbour users $Neighbour_{a,y}$ of active user a for target item y;

   It can be identified using 2 techniques $Neighbour_{a,y}$:

- "Thresholding": fixing an outright user closeness "threshold w", to know neighbours $Neighbour_{a,y}$ of active user with outright closeness $w_y(a,x)$ bigger than w.
- "Best-neighbours": selecting the the max neiNum similarity among $w_y(a,x)$ $(x \in 1,2,.,a-1, a+1,.,n)$, & corresponding neiNum users comprise $Neighbour_{a,y}$.

5. Mean of neighbours $Neighbour_{a,y}$ is used to calculate aweighted mean, prediction value $P_{a,y}$ is calculated of rating of user a for item y,

$$P_{a,y} = \overline{R_a} + k \sum_{x \in Neighbour_{a,y}} w_y(a,x)\left(R_{x,y} - \overline{R_x}\right),$$

$$\frac{1}{k} = \sum_{x \in Neighbour_{a,y}} w_y(a,x)$$

6. Most intriguing items of client an agreeing to $p_{a,y}$ $(y \in I_{P,a})$ are chose to produce recommendation to an active user. Following methodologies can be used

1. "Threshold": fixing an "rating threshold r", all prediction with value $p_{a,y}(y \in I_{P,a})$ are selected to make recommendation.
2. "Best-L-rating": choosing the maxi recomITEMNum predicting value from all $p_{a,y}(y \in I_{P,a})$ to make recommendation.

IV. CONCLUSION

The current thesis is concerned with Customer Targeted E-commerce & focuses mainly on Recommendation systems finding the recommendations based on purchases, the framework of E-commerce website has been outlined; the approach to the development of recommendation systems has been completely defined & delimited & has been broken into individual problems to be addressed.

\

REFERENCES

1. Zhang, Y. and Jiao, J. (2007). An associative classification-based recommendation system for personalization in B2C e-commerce applications. *Expert Systems with Applications*, 33(2), pp.357-367.

2. Li, Y., Lu, L. and Xuefeng, L. (2005). A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce. *Expert Systems with Applications*, 28(1), pp.67-77.

3. Huang, Z., Zeng, D. and Chen, H. (2007). A Comparison of Collaborative-Filtering Recommendation Algorithms for E-commerce. *IEEE Intelligent Systems*, 22(5), pp.68-78.

4. Simonetti Ibandntilde;ez, A. and Paredes Figueras, J. (2019). *Vuforia v1.5 SDK: Analysis and evaluation of capabilities*. [online] Upcommons.upc.edu. Available at: https://upcommons.upc.edu/handle/2099.1/17769 [Accessed 24 May 2019].

5. Sites.google.com. (2019). *2. Tutorials - MAMP & MAMP PRO Documentation*. [online] Available at: https://sites.google.com/site/mamppro/en/additional-resources/tutorials [Accessed 24 May 2019].

6. Ullman, L. (2012). *PHP and MySQL for Dynamic Web Sites: Visual QuickPro Guide, Fourth Edition*. [Berkeley, Calif]: [Peachpit Press].