

Collaborative Filtering Recommender System for a Website

Nishad Joseph Menezes, Vindhya Chandrasekharan,
V. I Eapen
Christ University Faculty Engineering
Bengaluru, India

Mausumi Goswami
Christ University Faculty Engineering
Bengaluru,
India

Abstract—The world wide web has ample amount of information. Not all of them seem relevant to a particular user. This is where recommender system defines its role. It allows the website to understand the user on a more personal level and recommends results that will interest them. There are mainly two vast subdivisions in the recommender system namely – Collaborative and Content based. Each finds its application on various website and our customized to the needs of a particular site. In this paper we will discuss collaborative an algorithm and share results.

Index Terms—Collaborative filtering, content filtering, e-commerce, recommender systems

1. INTRODUCTION

Recommender system has been used in almost every famous website. It is used to access query results that are mainly based on user interaction and searches on the website. It is a subclass of information filtering system that predict the rating and preference that each user give to a particular item. This system is used in several arena like movies, music, related articles social tags and search queries in general. Recommender systems are usually applied in three ways; collaborative filtering, content based filtering and hybrid that is a combination of both of these. In this paper we will concentrate on collaborative filtering and understand its subdivisions and grasp its concepts.

In this paper we will look at collaborative and content based filtering and in detail analyze the types of collaborative filtering that can be used on websites to better user engagement, we will also look at the algorithm, pseudo code and practical implementation of the same.

2. COLLABORATIVE FILTERING APPROACH

One of the approaches to recommender system is recommendation system. In this method, we collect and analyze user's behavior on a large scale. Their activities, preferences and likes are used to predict the likings of another user with similar results. The advantages of this system is that it does not depend on any machine analyzed content and only on user input and therefore can recommend results to complex systems such as shopping, movies or any other information in a social networking website. This recommender system is based on the assumption that if people have agreed before on a particular item, they will agree on another item again. The collection of data on collaboration filtering is either explicit or implicit. Explicit data collection includes asking the user for

his preference like asking to rate an item, to rank a collection or presenting two items and asking the user to choose a better one. Implicit data collection includes analyzing and observing the users view, keeping and tracking the records of the user's purchases and using cookies to track the user's behavior on the internet.

3. CONTENT BASED FILTERING

Content based filtering is based on description, the profile of the user and the user's preference. This recommender works more or less like a search engine that assigns a chunk of keywords to an item and uses the profile of the user to understand if the user would prefer the item. In short, it is based on what the user liked before and if the user would like something similar in the present. A lot of items are compared with previously liked items of the user and fitting items are recommended. Content based filtering works on information retrieval and information filtering. The method uses profile for an item and characterizes the item preferably using tags. It creates a weighted vector with every feature of the item and compares it to the likes and dislikes of the user analyzing the user's profile. The weights of the vector hire if it is an important feature from the user perspective. Later, all these rated contents are given appropriate preference using various techniques.

An issue that exists with content based filtering is that the system has to learn preferences of the user from his/her actions. When the system does not have enough items, then the system is significantly less than the other services that can be recommended.

4. ITEM BASED COLLABORATION FILTERING

The item based approach takes a set of items that a user has rated and then computes the similarity between target items and other items. Their corresponding similarities are also taken into consideration. When similar items are found, a weighted average is put on them. There are two ways of acquiring this, namely item similarity and prediction generation which we will discuss now.

A vital step in item based filtering is to understand the similarity between items and then to select the most similar item. Item similarity works as follows: two items A and B are first isolated and a similarity function $S(a,b)$ is found out. There are many ways to achieve this computation. Few of them are: Cosine based, correlation based and adjusted cosine based.

5. ITEM BASED COLLABORATIVE FILTERING ALGORITHM

In this section, we devise an algorithm for item based collaborative filtering. The algorithm scales to large data sets and produces acceptable recommendations in real time. Instead of matching a user to another similar user, this filtering matches each item liked to another item of similar features. It then combines all the similar items into a recommendation list. The simple algorithm for this filtering is given below:

Step 1: Users input a collection name

Step 2: Calculate the similarities, using

Step 3: The similar items are passed on as arguments in the recommendation function which considers the items liked and viewed and predicts the recommendation using the inputs from Step 2.

Step 4: A weighted average of all these recommendations is calculated and are displayed to the user.

Step 5: The final recommendation is displayed to the user based on their weighted average.

The recommendations are based on items preferred by the user and similar items and all related items. Below is a pseudo code for the same.

For each Item i liked

For every User U who liked Item i

For each Item J liked by User U

Record Item (i,j)

Compute the similarity between Item i and

Item j

Item to Item filtering is a technique where users are not compared with other users but their rated interest in items is calculated. In this approach, similarities between i and j are computed offline by a simple formula:

$$\frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Where n -number of samples; x and y are the two inputs in our

case the 2 ratings by different users, Now, the rating for item

for user a can be predicted using a simple weighted average, as in: Weighted Average = $[(n1 \times N1) + (n2 \times N2)] / (n1 + n2)$

where $n1, n2, \dots$ are relative weights and $N1, N2, \dots$ are values

6. EXPERIMENTS AND RESULTS

It is a model based algorithm. In this algorithm the similarities of various items on the datasets are calculated by taking in account the similarities between the items to recommend relevant information to the user.

The algorithm works in two ways:

1. Makes recommendations based on popular searches.
2. Makes recommendations based on rating of the collection

6.1 RECOMMENDATIONS BASED ON PREVIOUS SEARCHES:

For this the algorithm takes in two parameters, the name of the collections and the name of the person. The algorithm then sorts out all the collections based on ratings given by different users that the said user follows and displays them in descending order of their rating.

The algorithm first checks the current user with other users that the said user is following and calculates a similarity distance between the users. For each of the followed users, the algorithm checks if their preference has a collection in common. If there is a collection in common, it is excluded, else it finds the average of all the other containers and displays it in descending order of their rating.

Figure 1 shows the output for the aforementioned algorithm, using the input values from Table 1.

User	Collection	Rating
Phil	Best Coldplay songs	2.5
	Politics	1.0
	Arsenal	3.0
	Henry's best goals	4.0
	Best Cricketing moments	2.5
Sameer	Best Bikes of 2014	3.5
	La Liga's best moments	2.5
	Politics	1.0
	Best Cartoons of the 90's	3.5
	Henry's best goals	3.5
John	Best cricketing moments	3.5
	Best bikes of 2014	1.0
	Budget 2015	5
	Best Crime novels	3.5
	Arsenal	1.0
Peter	Top 10 movies 2014	5
	Php notes	3.5
Jill	La Liga's best moments	1.5
	Best Crime Novels	2.5
	Best Cartoons of the 90's	4
	Best comedians	3.5
	The world without end	2.5
	Best bikes of 2014	3.5
	La Liga's best moments	3
Bruce	Best of Jazz	1.5
	Best cartoons of the 90's	3.5
	Henry's best goals	3.5
	Tourist spots in India	2
	Best Bikes of 2014	3
Tom	Top 10 movies of 2014	2.5

Table 1: Input Data for Algorithm

Recommended Tag	Rating
Budget 2015	5
Henry's Best Goals	3.8041996461162
Best coldplay songs	2.5
Arsenal	2.33333333333
Tourists spots in india	2
Best of Jazz	1.5
Politics	1

Figure1: Displaying the array of preferred recommendation

7. CONCLUSION

In this paper we discuss the subdivisions of recommender systems. We further discussed collaborative filtering and their types. We then discoursed a generic algorithm and pseudocode for item based collaborative filtering. The implementation of this algorithm in php,scripting language and their results have been shown. The idea of the paper was to understand collaborative system and item based collaborative filtering in particular. A naïve user can just look upon the results and approve of the recommendations if item based is the recommender he/she prefers. Item based considers other users aspects also into its predictions while content based is confined to its own available information. Item based approach proved to be more efficient as compared to the content based approach and therefore was implemented, however depending on a user's need either of the recommendations prove to be useful. The output of this implementation can be used for the real time implementation of a recommender system on say, an e-commerce or a social networking website.

8. REFERENCES

- 1) J. Bobadilla, F. Serradilla and A. Hernando, 'Collaborative filtering adapted to recommender systems of e-learning', Knowledge-Based Systems, vol. 22, no. 4, pp. 261-265, 2009.
- 2) G. HUANG, 'Approach to collaborative filtering recommendation based on HMM', Journal of Computer Applications, vol. 28, no. 6, pp. 1601-1604, 2008..
- 3) Z. Huang, D. Zeng and H. Chen, 'A Comparison of Collaborative-Filtering Recommendation Algorithms for E-commerce', IEEE Intelligent Systems, vol. 22, no. 5, pp. 68-78, 2007..
- 4) C. Jensen, P. Pilkauskas and T. Lefèvre, 'Classification of Recommender Expertise in the Wikipedia Recommender System', Journal of Information Processing, vol. 19, pp. 345-363, 2011.
- 5) H. KumarKushwaha and J. Jeysree, 'Personalized Recommender System: A Personal Recommender System Online Social Networking Sites', International Journal of Computer Applications, vol. 93, no. 9, pp. 1-6, 2014.
- 6) Q. LIU, K. MENG and S. LI, 'statistical process control in collaborative recommender system', Journal of Computer Applications, vol. 32, no. 3, pp. 707-709, 2012.
- 7) W. Yuan, D. Guan, S. Lee and J. Wang, 'Skeleton Searching Strategy for Recommender Searching Mechanism of Trust-Aware Recommender Systems', The Computer Journal, 2014.

6.2 RECOMMENDATION BASED ON RATING OF THE COLLECTION:

For this, we use an inverted array in which each collection contains the value of the number of people subscribed to it. The algorithm then uses this array to recommend a collection. This is done in two steps using two functions:

1. TransformPreference Function
2. MatchItems Function

In the first step, the algorithm makes use of the TransformPreference function to invert the array. The TransformPreference function takes the name of the collection and finds all the users subscribed to it, and then creates an array for each collection.

In the second step, the algorithm makes use of matchItems function to match previous searches to recommend the highest rated collection.

Figure 2 shows the output for the best collection recommended with regard to a previous search query 'Henry's Best Goals', is the collection 'Politics', with a match probability of 28.89%.

Recommended Tag	Rating
Budget 2015	5
Henry's Best Goals	3.8041996461162
Best coldplay songs	2.5
Arsenal	2.33333333333
Tourists spots in india	2
Best of Jazz	1.5
Politics	1

The best match for the searched collection is *politics* with a similarity of 28.89%

Figure 2: Displaying the probability match