Comparison of Content Based and Collaborative Filtering in Recommendation Systems

Conference Paper · January 2021

CITATION
1

READS
2,756

2 authors:

Abdul Aziz

3 PUBLICATIONS **35** CITATIONS

SEE PROFILE

Mubashra Fayyaz

National University of Computer and Emerging Sciences

1 PUBLICATION **1** CITATION

SEE PROFILE

ISSN 1975-4736ⓒMITA2019

Comparison of Content Based and Collaborative Filtering in Recommendation Systems

Abdul Aziz
Department of Computer Science
National University of Computer & Emerging Sciences
Karachi, Pakistan
abdul.aziz@nu.edu.pk

Mubashra Fayyaz
Department of Computer Science
National University of Computer & Emerging Sciences
Karachi, Pakistan
mubashra.fayyaz@nu.edu.pk

***Abstract*—In today's technological era most of the usersare consistently searching many things on internet and for searching we usually use search engines. It's good to have most relevant results of searches, which could be achieved through Recommender systems. Recommender systems makes easy for users to find and evaluate items of their interest from the bundles of choices. Essence of Recommender systems is to calculate the connection among the products and users based on user's preferences. In our paper, we present the difference between the two main approaches of recommendation systems one is Collaborative Filtering other is Content based Filtering.**

Keywords—Recommendation Systems; Collaborative Filtering; Content Based Filtering;

I. INTRODUCTION

Recommender systems are the branch of information filtering systems. These are the kind of systems which help in extracting the resembling things. Whenever you intend to buy something online for example on Amazon, Netflix or on Facebook. Same story runs on the backend of every online store. The mechanism that runs behind surprisingly helpswebsites toincreasetheir business revenue. These kinds of systems for online purchasing systems leads to somewhere between 10 to 25% increase revenue. The need to develop this kind of system is because there was a gap between information collection and information analysis by filtering all the available information to present what is the most important and valuable to the user. Recommendation systems are considered to be beneficial to both users and service providers. The main reason behind that that kind of systems reducetransaction costs of findingandselectingitemsin anonlineshopping environment. The two basic kinds of algorithms are usedwhen we talked about generating recommendations Content- based filtering and Collaborative filtering.

Content-based filtering approaches:

Overall basis of this algorithm is on similarities between items, items could be movies, songs, books, or two purchased items. This algorithm works on a series of distinct features of an item in order to suggest new items with similar kind of characteristics. CBF is generating user's preferences with respect to given data.

Collaborative filtering approache:

Collaborative filtering approach works on the basis of user's past orientation and then build a model about the items that has been previously purchased/selected by user or provided any numerical ratings to those items as well as related decisions taken by the other customers. Model then predicts the items that the user may have concern with.

II. COMPARING THE TWO APPROACHES

A. CONTENT-BASED FILTERING:

Content-based filtering is also called cognitive filtering. Content based filtering algorithms are based on the contents of an item/product and a list of the user's preferences which shows kinds of items user may like. Therefore these algorithms Recommends things which are similar to the things user liked before. The description of each object is expressed as a group of descriptors. The profile of user is shown with the similar terms and it is constructed by figuring out the content of things seen by the user. Recommendations of new item would be based on (previously) positive rated items. Algorithm ofcontent based filtering involves two tasks: information retrieval and information filtering. Resemblance can be figured out using several kinds of models, in content-based filtering to produce suggestions that are valuable for the user. Let's suppose, a user visits an online store an online store and searches for "Rimmel", "MAC" and "Loreal" the system will suggest products related to cosmetics and healthcare range. (Aggarwal, 2017)[1].

Unlike Collaborative based methodology content based has nothing to do with the user or the user's profile because it does not affect its suggestion process in any way.

Profile of user's preferences and item descriptions play an essential role in CBF. Recommendations of items in content-based filtering algorithms is by counting resemblance. The most excellent relating objects are suggested by examining items that the user rated earlier anytime with possible candidate items. A content based recommender performs its recommendations through the data provided by users, data could be in form of rating (explicitly) or in form of clicking on a link (implicitly). User profile is generated on the basis of data which is use to make future recommendations to user.

The accuracy in model is depend on more user inputs and taking actions on recommendations.

Figure1 shows the process how content based works. Which includes the following steps:

1.Extract the features of products for suggestions.

2.Perform comparison between the orientations of the users and the attributes of items.

3.Suggest products on the basis of the characteristics which fulfills the user's interests. Possible that depicts all the components of the system.

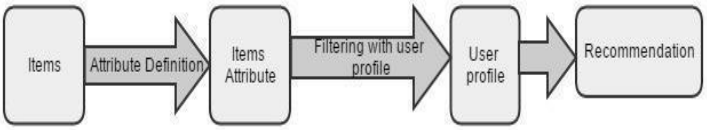


Figure 1 Content–based filtering

Term Frequency (TF) and Inverse Document Frequency (IDF), these two concepts are used for information extracting systems and content based recommendations. Term Frequency is the frequency of a term in a file. Inverse Document Frequency is the inverse of the document frequency in the whole file. The TF-IDF representation is most widely used algorithm (also called vector space representation). TF-IDF weighting contradicts the results of huge occurrence terms in defining the significance of an element. On the other hand when we compute TF-IDF, log is used to stifle the outcome of great occurrence terms. "For example: TF = 3 vs. TF = 4 is immensely different from TF = 10 vs., TF = 1000."[1] The conformance of a term during a document cannot be measured as an easy raw count and therefore the equation is below:

Equation:

$$w_{t,d} = \begin{cases} 1 + \log_{10} \mathrm{tf}_{t,d}, & \text{if } \mathrm{tf}_{t,d} > 0 \\ 0, & \text{otherwise} \end{cases}$$

Term Frequency	Weighted Term Frequency
0	0
10	2
1000	4

Figure 2 Formula of CBF

This is observed that the impact of parts that were in huge occurrence is now less intense and these computed answers are additional corresponding to one another as against the fi initial word occurrence. The similarity of items to one another is determined with the help of Vector Space Model it calculates the nearness on the basis of the angle in the middle of vectors.

According to this prototype, every piece is gathered in the form of vector from all aspects that are a part of n-dimensional space and the angles in the middle of the vectors are calculated using the correlation in the middle of the vectors.

Now, user preferences vectors are generated with the support of user's activities that are tracked earlier aspects of things and therefor the correlation in the middle of an element and a user is also resulted in a very comparable way.

1.Logarithmic inverse of the document frequency is taken to compute IDF surrounded by the complete frame of files.

Let's suppose, we searched for some word and the resulted set of documents constituted of 1 million documents out of which the required word is present in 0.5 million documents only so the IDF score computed to Log10(10^6/500000) that is equal to 0.30.

2.By taking square root of addition of the squared values of every feature during vector is the length of the vectors.

3.After calculating the TF-IDF weights and the size of the vectors, next step is to normalize the vectors through perform division on each vector by the text vector length.

4.Forth step, is to compute cosine values to compute the correlations.

For computing Cosine values– Assumed two documents pj and pk are shown as vectors of weight and their correlation

$$\mathrm{Sim}(p_j, p_k) = \frac{\overrightarrow{p_j} \cdot \overrightarrow{p_k}}{|\overrightarrow{p_j}| . |\overrightarrow{p_k}|} = \frac{\sum_{i=1}^{n} w_{i,j} \cdot w_{i,k}}{\sqrt{\sum_{i=1}^{n} w_{i,j}^2} \sqrt{\sum_{i=1}^{n} w_{i,k}^2}}$$

can be computed as

Figure 3 Cosine Formula

Where:

Wi, j = Weight of word I in document j Wi, k = Weight of word in document k.

Through this concept one can easily find out that which document user will like the most from the n numbers of documents.

Advantages of content-based filtering

1.Provides user freedom through ratings which are used to build user's profile.

2.It provides Transparency to its users by giving explanation how recommender system works.

3.Are advantageous for new users because they are enough to commend items that are which are not even located by single user.

4.It is not essential to have huge user set to attain the accuracy in recommendations.

5.New items can be commending instantly if item contents are available

Restrictions of Content Based filtering

1.It is a hard chore for CBF to produce the qualities for items in some regions.

2.Content Based filtering commends same sorts of objects for the reason that it bears from an overspecialization problem.

3.Content Based filtering are actually not good in grasping inter-dependencies or complicated actions.

Just in case if someone might like courses of Data Sciences, which includes real-world implementations beside with the concepts, but not only a concepts. Such kind of material cannot be seized by this kind of recommendation applications.

B. COLLABORATIVE FILTERING:

Collaborative filtering is also called social filtering in which other user's recommendation is used to filter information. The basic idea behind collaborative Filtering is assuming that if previously users showed agreement on some product would probably agree in future too. There are two approaches for Collaborative filtering:

i. User-based Collaborative Filtering (UBCF):

User is the main guy in user-based collaborative filtering. Grouping is done among users on the basis of their taste, interest and behavior. Within a group, every user gets recommendations of items evaluated by any member of same group depending on the preferences they share. InUBCF, similarity is calculated among users by KNN algorithm. User profile, interests and behavior are the features to find out the nearest neighbor. An example scenario is illustrated in Fig 3a. User1 and User3 both accessed a document A. Our system analyzes profile of both users as similar. Afterwards user3 accesses another document B. As both the users were marked similar, so B was also recommended to user1, assuming he will like it also.

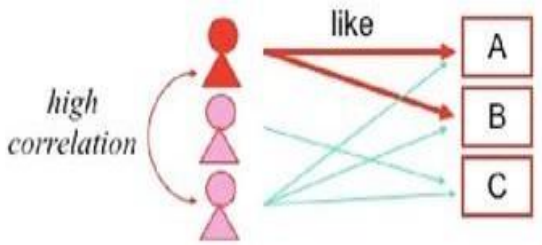


Figure 4a: User-based collaborative filtering.

User-based CF includes the following steps:

1.Users whom have the identical rating design with the dynamic user are found.

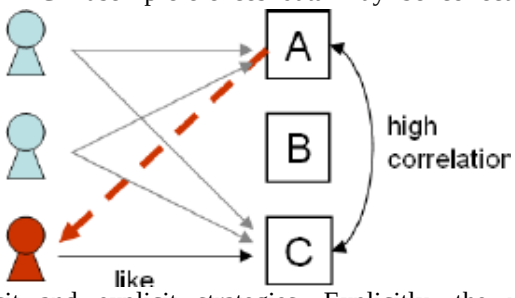
2.Prediction for the active user is done on the basis of the ratings of those compatible/similar users.

ii. Item-based Collaborative Filtering (IBCF):

In item-based collaborative filtering the similarity of the products is computed then on the basis of highest similarity products are predicted. After calculating the similarities among the products, a set of similar products is formed for the targeted user. Let's see an example scenario illustrated in Fig 3b, where items A B and C are analyzed and checked for similarity, Item A is similar to item C. Now if the user shows interest in item C. He would be recommended item A too.

Figure 4b: Item-based collaborative filtering.

In IBCF user preferences data may be collected using



implicit and explicit strategies. Explicitly, the user can practically select a rating score for an item on a certain statistical scale whereas implicitly, the users buying record or by clicking through rate may be analyzed. [4].

Item-based CF includes the following steps:

1.Relationship between items is represented through an item-item matrix.

2.The item-item matrix and user's data are matched and the taste of current user is predicted.

Benefits of Collaborative filtering

Using Collaborative filtering techniques which are based on memory

1.Implementation of recommendation system is made easier.

2.New data can be added easily by anyone and in incremental manner.

3.Prediction performance is improved.

Limitation of Collaborative filtering

1.For user to make exact recommendations CF systems need a large bulk of data that is already present.

2.We live in a world of so many devices and gadgets that produce millions of data every second so to work on such a large amount of data and make recommendations our system requires very large computing ability.

3.Due to the large amount of items that run out of stock in any online store so the rated items are a few in number.

S#	Difference	Content Based filtering	Collaborative Filtering
1	Background	Properties of items in I.	Ratings from users of items in I.
2	Input	User's ratings of items in I.	Ratings from user of items in I.
3	Process	Produce a classifier that fits user's rating preferences and use it on i.	Indicate users in U similar to u, and assume from their ratings one.

Table. 1. Comparison between CBF and CF

III. CONCLUSION

Recommender systems are becoming an advantageous tool that will generate suggestions to its users

according to their preferences. For the improvement of the suggestions accurateness in recommendation systems mechanism of purifying is used. Content-based recommendation systems suggests an item to a user based on some features of the item and a profile of the user's preferences. User itself maintains his/her profile, it is generally accomplished by the comments or ratings the user provides to the items. Most of the learning algorithms accommodates to learn user's profiles, and kind of learning algorithm is depending upon the representation of content.

Collaborative filtering algorithms practices "User Preferences" to suggest items. CF utilize actions of its further users and objects in form of different parameters like purchasing behavior, ratings, choice and purchasing descriptions. User's performance and orientations over the objects are helpful to suggest objects to its fresh users.

IV. ACKNOWLEDGMENT

We are highly thankful to Prof. **Kwang-Man KO** from Sangji University, South Korea and **Dr. Fahad Samad** from National University of Computer & Emerging Sciences, Karachi Pakistan, for theirguidance and encouragement of this research work, without thier kind guidance and help this work would not be completed.

References

[1] bhavya sanghavi, r. R. (2014). Recommender system- comparision of content based filtering and colaborative filtering . International journal of current engineering and technology.

[2] Aggarwal, p. (2017). Comparing content based and collaborative filtering in recomendation systems. International journal of new technology and research, 65-67.

[3] Ashish pal, p. P. (august 2017). An improved content based collaborative filtering algorithm for movie recommendation. International conference on contemporary computing, 10-12.

[4] Donghui wang, y. L. (may 2018). A content-based recommender system for computer science publications. Knowledge-based systems.

[5] Jian wei, j. H. (2016). Collaborative filtering and deep learning based recommendation system for cold start items.

[6] K. Heung-nam, e. A.-s. (2017). Collaborative error- reflected models for cold-start recommender systems. Decision support systems , 519–531.

[7] Mehrbakhsh nilashi, o. I. (2017). A recommender system based on collaborative filtering using ontology and dimensionality reduction technique. In expert systems with applications.

[8] Poonam b. Thorat, r. M. (january 2015). Survey on collaborative filtering, content-based filtering and hybrid recommendation system. International journal of computer applications, (0975 – 8887) volume 110– no. 4.

[9] Y., 4. L. (2015). Recommendation system using collaborative filtering. Master's projects, 439.

[10] Zhongqi luy, z. D. (2015). Content-based collaborative filtering for news topicrecommendation. Aai conference on artificial intelligence.

[11] Survey on collaborative filtering, content- based filtering and hybrid recommendation system. International journal of computer applications, (0975 – 8887) volume 110 – no. 4.

[12] Zhongqi luy, z. D. (2015). Content-based collaborative filtering for news topic recommendation. Aai conference on artificial intelligence.