International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 - 5161

Available at http://inpressco.com/category/ijcet

General Article

Recommender Systems - Comparison of Content-based Filtering and Collaborative Filtering

Bhavya Sanghavi[Å*], Rishabh Rathod[Å] and Dharmeshkumar Mistry[Å]

[Å]Computer Science, Dwarkadas J.Sanghvi College of Engineering, Vile Parle(W),Mumbai-400056,India.

Accepted 05 Sept 2014, Available online 01 Oct 2014, **Vol.4, No.5 (Oct 2014)**

Abstract

Recommender systems is a computer-based method that helps the user by generating suggestions about new items and products. It does so with the help of the past ratings of the item or analyzing the preferences of the user's friends in the social network. The recommender system is further optimized by considering the user demographics which further help in filtering the output. Recommender systems have a wide range of applications. It ranges from movies music, news, books, products to research articles, search queries, social tags, etc. The recommender system uses collaborative filtering algorithms. The two techniques are content-based filtering and collaborative filtering. In this paper, we focus on the positives and negatives of both the techniques.

Keywords: *Recommender Systems, Collaborative Filtering, Content-based Filtering, Hybrid Filtering.*

1. Introduction

The need of the day in the 21st century on the social media platform has been analyzing user preferences. It has been the fulcrum of every social media organization to develop filtering algorithms which scrutinizes the user data to generate implicit user preferences. With tremendous amount of data available, the algorithms focus on deriving relevant data which can be used by the users in their future endeavors on the social media platform. For example, the algorithms help the user by suggesting movies based on user's past preferences or user's social network in case of user wanting to watch a movie but is unsure of which movie. The filtering algorithms, thus, assists the user in making choices. These algorithms are classified as content-based filtering algorithm and collaborative filtering algorithm based on the approaches they use.

Content-based filtering compares the profile of a specific user with the items that are analyzed. Each item consists of descriptors or words that are related to the item. The user profile also consists of a set of descriptors which are added depending on the items previously seen by the user. Thus as the user explores various items, the user profile is enriched with strategic information which is then used by the learning algorithm of content-based filtering.

Disadvantage:-The key point to be noted with content-based filtering is whether the learning algorithm can predict about other categories of items with the help of user preferences for another category.

Example:- Depending on the history of the user with respect to a particular item, say the type of restaurant, the recommender system using the content based filtering can suggest another restaurant serving the same type of cuisine assuming that the user would like it. But having the same knowledge, the content based filtering cannot be used to suggest music to the same user. This is because there is a shift in the item category whereas the user preferences are limited only to the previous type of category.

Collaborative filtering works on the assumption that if a set of users have agreed on a certain idea at some point of time, it is very likely that they would agree on another related idea in future. This approach is dependent on the area of interest where the recommendation systems function.

Disadvantage: - Though this approach works pretty well using algorithms like KNN and Pearson Correlation, the main requirement of this approach is large amount of user activities and preferences. Since such data is not available since the activation of user profile, methods like cold starts are used that assumes user's preferences initially and then as the user provides preferences indirectly, it is implicitly collected and used to improve the accuracy of the recommender system.

Example:- If a set of users have liked a particular movie, say X-men, a particular user belonging to that set is suggested movies liked by other users of the respective set only.

2. Collaborative Filtering

Most of the algorithms using this approach use the nearest neighbor technique. The neighbors of a specific user is calculated by the proximity of the interests they share. Consider the below example. The columns indicate whether the user likes a particular movie or does not. A

*Corresponding author: **Bhavya Sanghavi**, **Dharmeshkumar Mistry** is working as Assistant Professor

'+' sign indicates that the user liked the movie where as the '-' indicates the opposite. Let us predict whether Ken will like the movie Frago. On examining the table, we realize that nearest neighbor of Ken is Mike since they have similar vectors due to their matching choice for other movies. Hence the recommender system suggests the movie to Ken.

	Amy	Jef	Mike	Chris	Ken
The Piano	-	-	+		+
Pulp Fiction	-	+	+	-	+
Clueless	+		-	+	-
Cliffhanger	-	-	+	-	+
Fargo	-	+	+	-	?

Fig.1 Matrix showing likes and dislikes of users

The problem with this method is high reliance on a single user which is not recommended. Hence, instead of depending on a single user, we take weighted averages of all the users in the neighborhood. This method not only increases the accuracy but also decreases the probability of the recommender system going off track. The following formula is used to calculate the co-relation between Ken and each of the other users. This co-relation r(X,Y) acts as a weight while predicting whether Ken would like Frago or not. X(k) and Y(k) are the ratings of user X and Y for item K, and $\bar{X}$ and $\bar{Y}$ indicate the mean of the ratings of the user.

$$r(X,Y)=\frac{\sum_k (X_k-\bar{X})(Y_k-\bar{Y})}{\sqrt{\sum_k (X_k-\bar{X})^2 \sum_k (Y_k-\bar{Y})^2}}$$

Now to predict the rating of Ken, we use this co-realtion of Ken with his neighbors and calculate using the following formula

$$p(X_i)=\frac{\sum_Y Y_i \cdot r(X,Y)}{n}.$$

Depending on this result, the recommender system might or might not recommend the movie Frago to Ken.

3. Content-based Filtering

Content based filtering algorithms are based on the description of an item and an offhand list of the user's preferences indicating a type of item the user likes. Thus, these algorithms suggest items that are similar to the items which are liked by the user in the past. This algorithm involves two tasks: information retrieval and information filtering. The information which includes the features of an item is extracted with the help of vector spacing representation. This information is developed into a user profile by considering the user's preference and the user's interaction with the recommender system. It creates a content-based profile of users based on the weighted vector of item features. Different techniques that can be used for the profiling include Bayesian classifier, cluster analysis, decision trees, artificial neural networks, etc.

Table 1 Comparison based on various parameters

Technique	Background	Input	Process
Collaborative	Ratings from **U** of items in **I**.	Ratings from **u** of items in **I**.	Identify users in **U** similar to **u**, and extrapolate from their ratings of **i**.
Content-based	Features of items in **I**.	**u**'s ratings of items in **I**.	Generate a classifier that fits **u**'s rating behaviour and use it on **i**.

4. Hybrid Filtering

To overcome the inaccuracies of both content based and collaborative filtering, a hybrid approach can be used. In this approach both the approaches for recommender systems can be applied independently and then can be added in an appropriate manner. Sometimes a model is used which incorporates features of both content based and collaborative filtering. Various techniques used in hybrid filtering include:

Weight - A score is given to each approach and then the results are combined.

Cascade - Each approach is given a certain priority before adding.

5. Future Work

Recommender systems are among the most profitable applications of data science in the big data world. Training data corresponding to the historical search, browse, purchase, and customer feedback patterns of your customers can be converted into golden opportunities for ROI. The predictive analytics tools of data science yield a bonanza of mechanisms to engage the customers and enrich the customer experience.

As we can see from the development of recommenders over the past years, the given results tend to be more personalised and subjective. Only considering the items itself and human ratings are no longer sufficient. A great amount of work in recent years have been done in music perception, psychology, neuroscience and sport which study the relationship between human preferences and the impact of human behaviour. The research now aims at targeting the human mood and thus, evaluating the user preferences based on the human mood.

For example, researches in psychology pointed out that music not only improves mood, increases activation, visual and auditory imagery, but also recalls of associated films or music videos and relieves stress. Moreover, the empirical experiments in sport mentioned that the main benefits for listening to the music which include work output extension, performance enhancement, and dissociation from unpleasant feelings etc . For example,

athletes prefer up-tempo, conventional, intense, rebellious, energetic, and rhythmic music rather than reflective and complex music. An important fact found by psychologists is that users' preference in music is linked to their personality. Also worth mentioning that fast, upbeat music produces a stimulative effect whereas slow, while soft music produces a sedative effects. All of these highlight that music recommender is not only a tool for relaxing, but also acts as an effective tool to meet our needs under different contexts. To our knowledge, there is few research based on these empirical results. Designing a personalized music recommender is complicated, and it is challenging to thoroughly understand the users' needs and meet their requirements.

Conclusion

As discussed above, the future research direction will be mainly focused on user centric music recommender systems. A survey among athletes showed practitioners in sport and exercise environments tend to select music in a rather arbitrary manner without full consideration of its motivational characteristics. Therefore, future music recommender should be able to lead the users reasonably choose music. To the end, we are hoping that through this study we can build the bridge among isolated research in all the other disciplines. Thus, the aim is to integrate the involvement of human mood in deciding the user preferences. It will help to increase the accuracy of recommender systems and to generate more reliable results.

Acknowledgment

We would like to thank our honorable principal Dr. Hari Vasudevan of D.J.Sanghvi College of Engineering and Head of Department of Computer Engineering, Dr.Narendra Shekhokar for giving us the facilities and providing us with a propitious environment for working in college. We would also like to thank S.V.K.M. for encouraging us in such co-curricular activities.

References

Joonseok Lee, Mingxuan Sun, Guy Lebanon (2012) A Comparative Study of Collaborative Filtering Algorithms, http://arxiv.org/pdf/1205.3193.pdf

Research Paper Recommender Systems: A Subspace Clustering Approach http://www.public.asu.edu/~ huanliu/papers/waim05.pdf

Item-Based Collaborative Filtering Recommendation Algorithms, http://files.grouplens .org/papers/ www10_sarwar. pdf

Collaborative Filtering Recommender Systems By Michael D. Ekstrand, John T. Riedl and Joseph A. Konstan, http://files.grouplens. org/papers/FnT%20 CF%20Recsys%20Survey.pdf

http://recommender-systems.org/collaborative-filtering/

http://recommender-systems.org/content-based-filtering/