Summary of Recommendation Algorithm Research

Jie Yan1, QingTao Zeng2* and FuQian Zhang3
1 Beijing Institute of Graphic Communication, BeiJing, China
2 Beijing Institute of Graphic Communication, BeiJing, China
3 Beijing Institute of Graphic Communication, BeiJing, China
*Corresponding author’s e-mail: zengqingtao@bigc.edu.cn

Abstract. In order to solve the problem of information overload caused by the rapid expansion of information, it is particularly difficult for users to select the information they are interested in when faced with massive amounts of information. Traditional search has some drawbacks and cannot solve the problem of information overload. The recommendation algorithm provided by the recommendation system is a good strategy to deal with. This article mainly introduces two recommendation algorithms and their principles, content-based recommendation algorithms and collaborative filtering-based recommendation algorithms. And their advantages and disadvantages are compared.

1. Introduction
In recent years, with the rapid development of the Internet era, the number of e-commerce companies has exploded, but the progress of the Internet seems to be a double-edged sword. While bringing great convenience to people’s lives, it also brings information overload. People have too many choices when choosing products. It is a good thing to choose, but too many choices are not a good thing. People pay more attention to efficiency in selection. The push of product information determines the success of e-commerce. The recommendation system uses recommendation algorithms to mine products that users are interested in to make relevant recommendations. At present, personalized recommendations are widely used in books (Dangdang), e-commerce (Taobao) and other fields. Now the more commonly used are divided into content-based recommendation algorithms, collaborative filtering-based algorithms, and hybrid recommendation algorithms.

2. The development of recommendation algorithms at home and abroad
During 1995, Lang et al. proposed a content-based recommendation algorithm, which extracts some attribute information of products that users have contacted, such as product label information, product introduction, and product evaluation. [1] During 1999, with the improvement of the accuracy of product recommendation, considering that the idea of similar interests of users in the same community was introduced into user recommendation, Pazzani proposed a recommendation based on statistical filtering, (DemographicFiltering, DF) algorithm. [2] Through the improved FOF theory, Naruchitparames et al. comprehensively consider the mutual friend information between users, and use the Pareto optimal genetic algorithm to predict the user's friend link and optimize the user's friend recommendation list. [3] By analyzing the topological structure of different user relationships in social platforms, Armentano et al. recommended users with high similarity within a certain distance to target users. [4,5] Akbari et al. not only considered the topological structure of users’ social relationships, but
also combined the FOF theory, and proposed an improved friend recommendation system based on topology and FOF theory. [6]

3. Introduction to recommendation algorithm

3.1. Content-based recommendation algorithm
First, extract the attributes that represent them from the item, and secondly, describe the user's behavior and habits through the user's past preferences. For example, if a user likes to read articles of the recommendation system type, then its favorite feature is "recommendation" and other keywords. Recommend related articles based on the cosine similarity between other items and favorite features.

3.2. Recommendation algorithm based on collaborative filtering
The collaborative filtering algorithm is mainly to collect answers from a large number of people and a large amount of data, to help people obtain a result in a statistical sense, and extract the common part from the large amount of data as a recommendation transmission. For example, the user wants to listen to a song. You will get the recommended playlist from users who have listened to the same song or friends with similar tastes through the same behavior as you did before.

3.2.1. The user-based collaborative filtering algorithm is to find the "neighbors of the user", and first find those users who have similar interests and preferences with the target user. The similarity here introduces the concept of Cosine similarity. We use the Euclidean distance to calculate the similarity, and the distance is the closest The user of is defined as the most similar user neighbor, and the closest item is defined as the most similar item neighbor. The main implementation steps of the algorithm here are divided into three steps. Step 1: Collect user information, clean and store data. Step 2: Calculate similar neighbors based on Euclidean distance. Step 3: Generate a recommended list through similar neighbors.

3.2.2. Item-based collaborative filtering is similar to user-based collaborative filtering. It finds similar items, obtains item information through the user's historical behavior records, and then calculates item cosine similarity to find neighbors of items to generate a recommended list.

4. Advantages and disadvantages of recommendation algorithm

4.1. Based on the advantages and disadvantages of content recommendation algorithm
Advantages 1. The advantages of content-based recommendation algorithms are the independence between users. The content recommended according to the preferences of each user is naturally irrelevant to others, but the recommendation algorithm based on collaborative filtering is on the contrary, recommendation based on collaborative filtering The method needs to obtain the favorite characteristics from a large number of users and then recommend them. Based on the content-based recommendation, the preference characteristics of others will not affect you.

Advantage 2. New content will be recommended immediately, if it is a purely collaborative filtering-based recommendation system, new items added will not be recommended

Disadvantages 1. The independence of users of content-based recommendation algorithm is also its disadvantage, and it cannot dig out the potential interests of users.

Disadvantage 2. Because the new user does not have a preference history record, there is no way to get the user's preference characteristics, so it is impossible to recommend him.

4.2. Advantages and disadvantages of recommendation algorithm based on collaborative filtering
Advantages 1. Randomness. A classic example. People who buy beer in the store are more inclined to buy diapers at the same time. Putting beer and diapers together increases the sales of the two products.
The content-based recommendation algorithm is for those who buy beer. Users will only recommend different brands of beer and will not recommend diapers, which is not random.

Disadvantages 1. When the product is just launched and when a new user just joins, it is impossible to predict the preference characteristics without the user's behavior data, so we have no way to recommend him.

Disadvantage 2. The sparseness of the data matrix. This problem occurs when users only rate a small part of the item. The larger the data scale, the sparser data matrix.

5. The basic principle of recommendation algorithm
Basic principles of content-based recommendation algorithm is as follows

![Figure 1. The basic principle of content-based recommendation algorithm.](image)

Suppose that the type of movie A that user A likes is romantic love, the type of movie B that user B likes is horror and thriller, the type of movie B that user C likes is horror and thriller, and the type of movie C is romantic love. Because the types of movie A and movie C are similar both are romantic love. Since user A likes movie A, he will recommend movie C of the same type to user A.

The basic principle of the recommendation algorithm based on user-based collaborative filtering is as follows.
Figure 2. The basic principle of the recommendation algorithm based on user-based collaborative filtering.

Because user A likes item A and item C, user B likes item B, and user C likes item A, item C, and item D. Since both user A and user C like item A and item C, user A and user C are similar users, so the item D that user C likes but user A does not like is recommended to user A.

The basic principles of item-based collaborative filtering recommendation is as follows.

Figure 3. The basic principles of item-based collaborative filtering recommendation.
User A likes item A and item C, user B likes item A, item B, and item C, and user C likes item A. Since item A and item C are similar, item C is recommended to user C who likes item A.

6. Existing problems and prospects for the future
(1) Recommendation quality and time efficiency. Recommendation algorithms often pursue recommendation accuracy and ignore time complexity. The complexity of recommendation algorithms is high, so we have to take some measures to optimize time complexity. Judge before the iteration, if its score is already very low, we judge that the interest shift has occurred, and we will directly eliminate it and no longer iterate to save time and efficiency.

(2) Regarding the issue of cold start, for new users and new projects, when a user has no behavior or an item has not been evaluated, it becomes difficult to recommend. The problem of sparse matrix is also involved here. We can consider adopting matrix singular value decomposition and clustering methods to alleviate this problem to a certain extent.

(3) Data cleaning issues. For some current malicious review issues and marketing account issues, we need to eliminate them to make better recommendations.

7. Conclusion
According to the above problems, I will use the clustering method to solve the cold start problem of new users or new projects. Since the time of graph traversal is particularly long, I will determine the PR value of each graph node to determine whether to continue the iteration. Reduce the impact on new projects and new users through half-life.

Acknowledgments
(1) Major Science and Technology Projects of Guangdong Province in 2019, No. 190826175545233
(2) Beijing science and technology innovation service capability construction project (PXM2016_014223_000025)
(3) BIGC Project (Ec202007)

References
[1] Lang K. NewsWeeder: learning to filter netnews[C]. San Francisco: Morgan Kaufmann Publishers Inc, 1995: 331-339.
[2] Pazzani M J. A Framework for Collaborative, Content-Based and Demographic Filtering[J]. Artificial Intelligence Review, 1999, 13(5-6): 393-408.
[3] Naruchitparames J, Güneș M H, Louis S J. Friend recommendations in social networks using genetic algorithms and network topology[C]. New jersey: Evolutionary Computation, 2011:2207-2214.
[4] Armentano M G, Godoy D, Amandi A. Towards a Followee Recommender System for Information Seeking Users in Twitter[C]. Berlin: Springer, 2013: 27-38.
[5] Armentano M G, Godoy D, Amandi A. Topology-Based Recommendation of Users in Micro-Blogging Communities[J]. Journal of Computer Science and Technology, 2012, 27(3):624-634.
[6] Akbari F, Tajfar A H, Nejad A F. Graph-Based Friend Recommendation in Social Networks Using Artificial Bee Colony[C]. Washington: IEEE Computer Society, 2013: 464-468.