Research on Application of Collaborative Filtering Algorithm in Digital Movie Recommendation

Sansan Li¹a*, Dongxian Zhou²

¹Department of Computer Science and Engineering, Guangzhou College of Technology and Business, Guangzhou, Guangdong, 510850, China
²Department of Economics and Trade, Guangzhou College of Technology and Business, Guangzhou, Guangdong, 510850, China

azhdongx@mail3.sysu.edu.cn  Co rresponding author’s e-mail: 542577885@qq.com

Abstract. Information overload will undoubtedly increase the time for users to search for information on the Internet. Data mining technology based on big data is constantly improving this situation. It extracts hidden, previously unknown information from massive data information to find out the internal connection between users, between the user and the item, and between items, so as to recommend information according to the user's preferences and needs, and even provide the user with a customized page. So how does it implement personalized recommendations? This article takes online digital movie recommendation as an example, and explains the principle and process of collaborative filtering recommendation algorithm to achieve personalized recommendation through the combination of theoretical analysis and experiment.

1. Introduction

The development of modern Internet technology has made the society usher in digitization, informationization and march towards intelligence. In this era of information overload, users need a certain period of time to find favorite information from a large amount of information, and even difficult to find favorite products or information, resulting in negative behavior of users. Think about it, where do we spend our time online? Studies have shown that the time users spend on the Internet is mainly used for social networking, reading, online shopping, various searches, emails, video sites, etc. The top three are: social networks account for 22%, and various searches account for 21%, Reading accounts for 20% [1]. It can be seen that obtaining information that meets our needs from massive data information is really not an easy task. In recent years, the rapid development of big data analysis and data mining technology has led to the emergence of a recommendation engine. It can accurately deliver information to users who need it, so that precise marketing and personalized services for various online platforms can be achieved. For example, Douban movies will obtain a comprehensive ranking of movies based on certain factors such as the number of people who have "watched", ratings, and evaluations. The Douban Guessing function of Douban reading is also based on data mining.

So in the era of big data, how does data mining technology achieve personalized recommendations? Data mining can use algorithms to extract hidden, previously unknown and potentially useful information from massive and diverse data sets [2]. Therefore, through data mining, you can obtain knowledge from the log records and user order information generated by the user's browsing page, find out the internal connections between items, between users, items and users, mine user needs and...
preferences, and then present information that may be of interest to users. Thus, it can shorten the time for users to find information and let customers experience more personalized services. Many e-commerce platforms now have personalized recommendation functions such as "guess what you like", "see similar" and "recommend for you".

2. Principle of personalized recommendation based on collaborative filtering algorithm

2.1 Personalized recommendation
Personalized recommendation is centered on user needs, and recommends information that may be of interest to users according to the user's browsing content, browsing habits and preferences, so that the user is no longer troubled by information overload. For example, when we go out to eat or watch a movie, we will face many choices. When we do not know how to choose, we may consult friends or service personnel around us to see what recommendations they have and then make a choice. This is a common example of personalized recommendation in life and the core idea of collaborative filtering.

In the field of e-commerce, there are many examples of personalized recommendations. Such as, Amazon judges customers' interests based on the products that users have marked "have bought, already owned, and rated", and then recommends products that may be of interest to users. eBay's newly launched "interest shopping" function can provide behavioral portraits for each shopper based on the shopper's browsing and shopping behavior, and then create a private customized page for the user. The key to personalized recommendation is to find the items of interest to the target user through a certain algorithm or to make recommendations based on the preferences of the user group with high similarity to the target user.

The implementation of the personalized recommendation function mainly depends on the recommendation engine. The recommendation engine is an information network which uses a variety of methods based on content, based on user behavior, based on social network, etc. to actively learn and mine the laws of the user’s current or potential needs, and push the products or content of interest or need to the user[3].

2.2 Collaborative filtering recommendation algorithm
The recommendation algorithm is the core of personalized recommendation. There are many algorithms that can implement personalized recommendation, such as: collaborative filtering recommendation algorithm, recommendation based on association rules, and content-based recommendation. Each recommendation method has its own advantages and disadvantages, and has its own application environment. This article mainly introduces collaborative filtering recommendation methods. Collaborative filtering algorithm (CF) can be said to be a kind of recommendation based on "collective collaboration", and its central idea is "Birds of a feather flock together". Normally, it can mine user preference information through user ratings, voting, forwarding, buying, etc., and then find items with higher scores from the user group with high similarity to the target user or find items similar to the items purchased or liked by the target user, and finally generate a recommendation list based on the predicted score. The implementation process of the collaborative filtering recommendation algorithm can be summarized as the following steps: data collection (data set), similarity calculation, determining the nearest neighbor, and generating a recommendation list [4], as shown in Figure 1 below:
Collaborative filtering recommendation algorithms can be divided into: user-based collaborative filtering algorithms and item-based collaborative filtering algorithms. The basic principle of collaborative filtering recommendation based on users is to find a set of users with high similarity to current user preferences, and find items that the user group likes and the target user has not acted on, and then recommend them to the target user. For example, if users A and B have watched three movies, movie1, movie2, and movie3, and give 5 points of praise. Then users A and B have a high degree of similarity. When recommending a movie to user B, we can recommend the movie4 that A has seen (user B has not seen) to him.

The item-based collaborative filtering recommendation algorithm is to find items similar to the items that the target user likes or has purchased based on the ratings of all users on the item, and then recommend it to the user. For example, if user A purchases Book 1 and Book 2 at the same time, it means that Book 1 and Book 2 are highly related. When user B buys book 1, we can infer that he also needs to buy book 2. Figures 2 and 3 give an explanation of the principles of the two recommendation algorithms.
3. Application of collaborative filtering recommendation algorithm in online digital movie recommendation

User-based collaborative filtering and item-based collaborative filtering have their own advantages in terms of diversity of recommendations and users' adaptability to recommendation algorithms. For example, for this experiment, the number of movies far exceeds the number of users, and movies are updated relatively frequently. Therefore, this article will use the user-based collaborative filtering algorithm to conduct experiment. The goal of the experiment is to find 20 similar users for each user and recommend 10 movies. The specific process is as follows:

3.1 Establish user-movie matrix model

Collecting users' preference information is the most basic content of personalized recommendation. The data set of this experiment is the official MovieLens data set. This experimental data set contains more than 100,000 ratings of 8,703 movies by 610 users, and the ratings of movies by users are 1-5. In order to better evaluate the model, we divide the data set into a training set and a test set with a ratio of 3:1.

The input data of the collaborative filtering algorithm is usually a user-item rating matrix, that is, R=M*N, M represents the user data set, N represents the movie data set, and Rij represents the user i's evaluation of the movie j. As shown in Figure 4.

![Figure 4. User-Movie Matrix](image)

3.2 Find a user set with similar interests to the target user

After analyzing user behaviors to get user preferences, we find similar users based on user preferences. Commonly used similarity calculation methods are: cosine similarity, Pearson similarity, and Euclidean distance. This paper uses cosine similarity to calculate the similarity between users. Find 20 users with similar interests to the target user. Calculated as follows:

$$w_{uv} = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)| \times |N(v)|}}$$

In formula (1), Wuv represents the similarity between user u and user v; N(u) represents the collection of items liked by user u; |N(u)| represents the number of items liked by user u. According to this formula, we can calculate the similarity between each user and the target user U, and then use the TOP-N method to rank these similarities and obtain the set S(u, 20) of the top 20 users with the most similar interests to user U.

3.3 Generate a recommendation list

Next, find the movies that they like among the 20 most similar users and remove the movies that the target user already likes, that is, extract the candidate item set. Finally, We need to calculate the predicted score of each candidate movie, that is, the degree of interest of the target user in the candidate movie, and recommend the higher-rated movie to the target user [5]. The recommended calculation formula is as follows:
In formula (2), $W_{uv}$ represents the similarity of interest between user $u$ and user $v$, $r_{vi}$ represents the interest of user $v$ in item $i$, $S(u,k)$ represents the K users closest to user $u$'s interest, $N(i)$ means a collection of users who like item $i$\textsuperscript{[6]}. In this article, $r_{vi}$ refers to the user's rating of the movie. Use this formula to calculate the predicted scores and rank them, and recommend the top 10 movies to the user. For example, the information of 10 movies recommended for User 1 is shown in Table 1 below.

### Table 1. User 1's movie recommendation list

| user | Recomended Movie ID | predictive score |
|------|---------------------|-----------------|
|      | 589                 | 3.84            |
|      | 1527                | 3.81            |
|      | 1200                | 3.55            |
|      | 608                 | 3.40            |
|      | 1968                | 3.34            |
|      | 1610                | 3.30            |
|      | 480                 | 3.18            |
|      | 3793                | 3.14            |
|      | 1036                | 3.12            |
|      | 858                 | 3.12            |

### 3.4 Evaluate the algorithm

After making predictions, this paper mainly evaluates the accuracy of the recommendation algorithm from the accuracy rate and the recall rate, and evaluates the algorithm's ability to discover the long tail or understand the diversity of recommended content through coverage. After calculation, the accuracy rate of this experiment is 29.38%, the recall rate is 17.16%, and the coverage rate is 4.24%. The experimental results show that the user-based collaborative filtering recommendation algorithm has some shortcomings in recommending long-tailed capabilities, and the accuracy of the experimental algorithm needs to be improved. Next, the recommendation quality can be continued by tuning and improving the recommendation algorithm parameters.

### 4. Future expectations

In the era of information explosion, how to provide users with more efficient and accurate information that meets their needs is an area worthy of study. Therefore, the research on personalized recommendation will continue to deepen with people's needs. With the development and maturity of big data and mining technology, the personalized recommendation is more inclined to hybrid recommendation algorithm, and continue to introduce deep learning technology. In this way, the recommendation engine can use context, social, visual, knowledge mapping and other more existing information, to achieve a more fine-grained extraction of the characteristics of users and items through learning, so as to know the users better than themselves.

### Acknowledgments

Supported by Foundation for Distinguished Young Talents in Higher Education of Guangdong, China(2015KQNCX194), and the quality engineering project "Research on Improving Teaching Ability of Applied University Teachers' Innovation and Entrepreneurship Education"(ZL20191130) of Guangzhou College of Technology and Business.

### References

[1] SocialBeta.(2012) [Infographic] Where do people spend their time when they go online?. https://socialbeta.com/t/online-time-spend.html.

[2] Shi, S.f. (2019) Big Data Analysis and Mining. People's Posts and Telecommunications Press, Beijing.

[3] Baidu Encyclopedia. (2020) Recommendation Engine.
[4] Liu, X.f. (2016) Research on Personalized Recommendation System Based on Web Data Mining. Changchun: Changchun University of Technology.

[5] Wang, M. (2019) Design and implementation of e-commerce personalized recommendation system based on collaborative filtering. Guangzhou: Guangzhou University.

[6] CSDN. (2017) Collaborative filtering recommendation algorithm. https://blog.csdn.net/congnaahahei/article/details/78940939.