

# Research on Commodity Recommendation Algorithm Based on Collaborative Filtering

Tong Haifeng<sup>1,2</sup>, Zhang Chengnian<sup>2</sup>, Hu Jianfeng<sup>2</sup>

<sup>1</sup> Jiangsu Business Intelligence Applied Engineering Technology Research and Development Center, Wuxi, Jiangsu, China

<sup>2</sup>Wuxi Vocational Institute of Commerce, Wuxi, Jiangsu, China

**Keywords:** Collaborative Filtering, User Characteristics, History Record, Products Featured, Algorithm

**Abstract:** the Rapid Development of e-Commerce Has Led to Tremendous Changes in the Retail Industry. More and More Offline Enterprises Have Begun to Turn to Online Development, and after Experiencing the Rapid Growth of Users, They Gradually Began to Face Problems Such as Slow Growth of Users and Disappearance of e-Commerce Dividends. At the Same Time, as Incomes Increase, Consumer Demand is Increasingly Diversified. Therefore, This Paper Constructs a Collaborative Filtering Algorithm to Study Commodity Recommendation, in Order to Solve the Development Problems Faced by Enterprises and Meet the Individual Needs of Users.

## 1. Introduction

### 1.1 Literature Review

To meet the increasingly diversified needs of consumers, it has become a focus area for retail companies and other companies. Commodity recommendations have entered the public's field of vision and received more scholars' attention. Yang Fang first introduced the traditional collaborative filtering algorithm, and analyzed the problems of recommendation algorithm in the perspective of user interest recommendation problem. A collaborative filtering algorithm based on user multi-interest was proposed, and this empirical evidence was found through empirical research. The method is effective (Yang et al., 2010) Wang Wei pointed out that the collaborative filtering algorithm is based on user behavior for personalized recommendation. The core idea is to recommend products based on the similarity of interests between users. Therefore, based on the traditional collaborative filtering algorithm, Wang Wei et al. used emotional online as an important influencing factor of user recognition, and studied the collaborative filtering product recommendation method. Experiments show that the user-based collaborative filtering algorithm based on emotional online has an effect on commodity recall rate indicators (Wang et al., 2014). Lu Chun et al pointed out that the collaborative filtering algorithm is one of the more effective algorithms in the recommendation system. The recommendation algorithm mainly scores the project based on other users with higher similarity to the user, and obtains a customer with higher similarity to the user (Lu et al., 2014). Liu Jingwei pointed out that in the era of rapid development of Internet shopping, providing quality goods and services has become the core content of major enterprises, and collaborative filtering algorithm is one of the more mature recommendation algorithms. Therefore, the collaborative filtering algorithm is analyzed in all aspects, and the collaborative filtering algorithm is proposed. Development direction (Liu, 2013).

### 1.2 Purpose of Research

With the rapid development of the Internet, e-commerce has gained a lot of room for development. In this context, more and more offline companies are beginning to switch from offline sales to online. More and more e-commerce companies are actively deploying offline sales strategies, and the boundaries between online and offline are becoming increasingly blurred. In such a big environment, Internet thinking and new technologies have begun to be widely used in

traditional sales fields to transform and innovate traditional marketing methods. Therefore, commodity recommendation is the main way to meet the increasingly diverse needs of users. The collaborative filtering algorithm is the most successful algorithm in the recommendation system, so it is of great practical significance to study the application of collaborative filtering algorithm in commodity recommendation.

## 2. Collaborative Filtering Algorithm

In the algorithm recommendation system, the collaborative filtering algorithm (UBCF) is the longest recommended algorithm in the application history (Yang et al., 2011). The specific calculation process of the collaborative filtering algorithm can be divided into the following steps. First, build the model. The user item scoring matrix is constructed according to the user network consumption behavior. By calculating the matrix, the intersection between the user and other user scoring items is obtained, and the user's intersection items are scored, and the user has the same point as other users. Second, find out the same set of neighbors as the user and make up the neighbor set (Gao et al., 2017). Generally, other users who have higher similarity with users can be determined by the following two methods. The first is a threshold for similarity presets. The similarity of other users to the target user, greater than this threshold, is the element of the neighbor set. The second method is to select a user with a higher similarity as a neighbor from the target user neighbors according to the similarity. The collaborative filtering algorithm mainly includes the following three types.

$$\text{Pearson correlation coefficient: } sim(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)^2 (r_{vi} - \bar{r}_v)^2}}$$

$$\text{Cosine similarity: } sim(u, v) = \cos(\bar{u}, \bar{v}) = \frac{\sum_{i \in I_{uv}} r_{ui} \cdot r_{vi}}{\sqrt{\sum_{i \in I_{uv}} r_{ui}^2} \cdot \sqrt{\sum_{i \in I_{uv}} r_{vi}^2}}$$

$$\text{Corrected cosine similarity: } sim(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u} (r_{ui} - \bar{r}_u)^2} \cdot \sqrt{\sum_{i \in I_v} (r_{vi} - \bar{r}_v)^2}}$$

Among them,  $u, v$  indicates the user,  $\bar{r}_v$  indicates the evaluation of all items of the product by the user  $v$ ,  $r_{vi}, r_{ui}$  indicates the evaluation of the  $i$  item by the user  $u, v$ , and  $I$  indicates the set of all the item combinations.  $\bar{u}, \bar{v}$  represents the scoring vector of  $u, v$ .

## 3. Model Building

### 3.1 Problem Description

At present, due to the rapid development of the Internet, more and more enterprises are beginning to develop and develop online retail. Therefore, it is necessary to understand the characteristics and information of user attributes, and then recommend different products to users to meet the individual needs of users. Therefore, in order to meet the individual needs of users, it is necessary to study the recommendation algorithm of commodities.

### 3.2 Problem Solved

User characteristics. that influence the choice of goods. There are often differences in the products that customers need, so user characteristics are one of the important criteria for classifying users. Users of the same category often have similarities in demand for goods. Therefore, when

performing product recommendation, detailed statistics of user characteristic information are required. User profile information mainly includes gender, age, address, and occupation category. In the specific operation process, the product recommendation system calls the user center system through the intermediate server to obtain the user data. Enterprises extract user data and classify users. In the specific classification, the ID of the store consumed by the user is also an important basis for classifying the user.

**Age.** At present, consumers at different ages have different lifestyles and a greater difference in demand for goods. For example, consumers in childhood are more concerned about snacks and toys. Consumers in adolescence are more concerned about beverages, daily necessities, etc. Consumers in the middle and old age are more inclined to buy some health products. Based on this, the paper divides users according to their ages. The age difference at each stage is 3 years old, that is, the age difference between the two users is less than 3 years old, and the two users are classified as one type of users. If the age difference between the two users is within 1 year, the similarity between the two users is higher. If the age difference between the two users is 3 years old, the similarity between the two users shows a gradual weakening trend.

**Gender.** Because of different genders, users also have large differences in demand for goods. Therefore, in order to make the product recommendation system better meet the individual needs of users, it is necessary to consider the gender of the user. In comparison, women's nature users buy cosmetics more than men. Therefore, when the product is recommended, a female user will recommend a large amount of cosmetics. Men are more likely to buy home appliances and hardware products.

**Occupation.** Different professional users also have large differences in the demand for goods. At the same time, in the product recommendation system, the customer occupation category has a greater impact on the recommendation system. Therefore, this paper makes assumptions when considering user occupation categories. Assuming that the two users have higher occupational similarities, the impact on the recommendation system is smaller. When the user's occupational similarity is low, the product recommendation system has a greater impact.

### 3.3 Model Building

Currently, the analysis features are mainly based on user similarity and historical scores. Therefore, based on user similarity and historical score, this paper further analyzes the user's final similarity. In this process, the user nearest neighbor matrix is constructed according to the final similarity of the user. In order to obtain the influencing factors of the commodity recommendation algorithm, the collaborative filtering algorithm is improved. Therefore, the similarity calculation is mainly performed for new users based on user characteristics. For the old customers, the historical score is used to calculate the similarity. The user eigenvalue and the historical score are calculated according to a certain weight ratio, and the calculation formula of the user similarity is finally obtained:

$$UserSim(x, y) = w \times ScoreSim(x, y) + (1 - w) \times AttSim(x, y)$$

$x, y$  is used to refer to the user,  $UserSim(x, y)$  is the final similarity between users,  $AttSim(x, y)$  is the user  $x, y$  feature similarity,  $ScoreSim(x, y)$  is the user  $x, y$  historical score similarity, and  $w$  is the user feature similarity and the historical score similarity weight. According to the calculation formula of the user similarity, the user similarity matrix  $\{U_1, U_2, U_3, U_4, \dots, U_n\}$  can be obtained, and  $S_{ij}$  is used to represent the similarity between the two users.

To this end, the paper further uses the K nearest neighbor algorithm to select neighboring K neighbors to further generate a matrix, as shown in table 1.

Table 1 User Neighbor Formation Matrix

User	$N_1$	$N_2$	...	$N_k$
$U_1$	$N_{11}$	$N_{21}$	...	$N_{k1}$
$U_2$	$N_{12}$	$N_{22}$	...	$N_{k2}$
...	...	...	...	...
$U_i$	$N_{1i}$	$N_{2i}$	...	$N_{ki}$
...	...	...	...	...
$U_n$	$N_{1n}$	$N_{2n}$	...	$N_{kn}$

### 3.4 Recommend

According to the neighbor matrix, the target user and its neighbors are tested, and the project is scored to obtain an average score. When the user is scored, the weighted average method is used to calculate the target user's predicted average score. The specific formula is as follows:

$$P_u = A_u + \frac{\sum_{n=1}^k (R_n - A_n) \times UserSim(n, u)}{\sum_{n=1}^k UserSim(n, u)}$$

Among them,  $UserSim(n, u)$  represents the similarity between users  $n$ 、 $u$ ,  $R_n$  represents the user  $u$  neighbor scores the project, and  $A_u$  represents the user  $u$  average score for the project. After obtaining the user's prediction of the item's prediction, the similarity may be sorted according to the score, and the corresponding item of the previous item is recommended to the user.

### 3.5 Algorithm Description

This paper first constructs a collaborative filtering algorithm model, and studies users based on user characteristics and historical records. Among them, the new user mainly adopts the user feature method to calculate the user similarity, and the old customer mainly studies according to the historical record. Finally, the user feature similarity and the historical score similarity are weighted to obtain the final similarity of the user.

### 3.6 Case Analysis

In order to verify the commodity recommendation algorithm under the collaborative filtering model, this paper analyzes the example. The data mainly comes from the Watsons data of large domestic retailers, mainly from the three aspects of user recall results, average absolute error, and coverage. The data is mainly based on part of the sales data from March to April 2019. The main source of access is the Internet. Among the collected data, 20001 of them are user purchase face data. To verify the above model, the user purchase data is divided into two categories. Among them, 80% of the user purchase records are used for test set A and 20% are used for test set B. Through experimental research, it is found that in the case of different neighbors, the products recommended by the users are different. The similarity between the recommended products and the neighbors with higher final similarity is higher. It can be seen that under the collaborative filtering algorithm, the product recommendation result is effective.

## 4. Conclusion

Research product recommendations by constructing a filtering synergy method. To do this, you need to calculate user similarity. When calculating user similarity, users are mainly divided into new users and old users. For the new user and the old user, the user matrix is calculated by using the user feature similarity and the historical score similarity respectively. On the basis of this, the weighted average method is used to calculate the final similarity of the user. The result shows that the product

recommendation accuracy is higher under the collaborative filtering algorithm.

### **Acknowledgement**

This Research has been financed by The Research and Development Center of Business Intelligence Applied Engineering Technology of Jiangsu Province in 2019 “Application of Collaborative Filtering Algorithms in Precision Marketing” (201908)

### **References**

- [1] Yang F., Pan Y.F., Li J., et al. (2010). An improved collaborative filtering recommendation algorithm. *Journal of Hebei University of Technology*, 39(3), 82-87.
- [2] Wang W., Wang H.W., Meng Y. (2014). Research on Collaborative Filtering Recommendation Algorithm: Consider Online Emotional Tendency. *Systems Engineering Theory and Practice*, 34(12), 3238-3249.
- [3] Lu C., Hong A.B., Gong J. (2014). Research on collaborative filtering recommendation algorithm based on PSO. *Computer Engineering and Applications*, 50(5), 101-107.
- [4] Liu J.Wi. (2013). Research on recommendation system algorithm based on collaborative filtering. *China Electronic Commerce*, 14(6), 157-157.
- [5] Yang S., Xue W., Xie Y.H., et al. (2011). Collaborative filtering recommendation algorithm based on single classification. *Computer Engineering*, 37(19), 59-61.
- [6] Gao C.Y, H.K., Wang J., et al. (2017). Collaborative filtering recommendation algorithm based on commodity attribute value and user characteristics. *Computer Engineering and Science*, 39(12), 2333-2339.