

# Agricultural Product Rating Prediction Design Based on Collaborative Filtering Algorithm

Bairui Tao, Ding Liu, Fengjuan Miao, Tongri Sun

College of Communications and Electronics Engineering, Qiqihar University, Qiqihar, Heilongjiang 161006, China

**Abstract.** The rapid development of agricultural e-commerce has led to more and more agricultural products on the website, and the problem of information overload has become increasingly prominent. By analyzing the user's historical score data set, this paper analyzes the recommended models of different collaborative filtering algorithms and different similarity calculation formulas, predicts the rating of the new agricultural products by the given user, tests the RMSE and MAE values under different algorithms, and helps the users to recommend the preferred agricultural products, thus providing personalized service to the users.

**Keywords:** agricultural e-commerce, collaborative filtering algorithms, rating prediction, RMSE and MAE.

## 1. Introduction

The electronic commerce of agricultural products is a new way for the sale and circulation of agricultural products, which is stimulated by the policy of hundreds of countries and the special funds of the Ministry of Commerce [1]. However, the present agricultural e-commerce platform and the mature e-commerce platform are compared horizontally, and there are many shortcomings in the present agricultural e-commerce platform. For example, the current e-commerce website is often a simple pile of agricultural products on the platform site, which leads to more and more goods on the website, and the problem of information overload is becoming more and more prominent [2,3]. For users, as the amount of information grows, a lot of information makes it difficult for customers to decide which farm products they are going to buy and overwhelm by this simple decision. The recommendation system provides users with suggestions to filter the amount of information that users need to process. In recommender systems, the most successful and widely used recommendation algorithm is collaborative filtering algorithm. Collaborative filtering analyzes user interests, finds similar users of designated users in the user group, and combines these similar users to evaluate a certain information, and forms a system to predict the preference of the specified user for this information. In the electronic commerce website of agricultural products, there are a large number of users' rating data for agricultural products. The rating data characterizes the user's clear view of the agricultural products, and has a high value for mining [4,5]. Through the analysis of the user's historical rating data set, the collaborative filtering algorithm is used to predict the given user's rating on the new agricultural products, to help the users to recommend the preferred agricultural products, and to provide personalized service to the users.

## 2. Rating Prediction

The rating is the user's grading of agricultural products after the purchase of agricultural products. The rating data are explicit feedback data, and the rating is predicted through collaborative filtering algorithm. The evaluation indexes are RMSE and MAE.

### 2.1 Predicting Process

Python is used to capture the rating data of agricultural products by the Jingdong, through the sorting and cleaning of data into a structured data table, divided into training set and test set according to rules, using the collaborative filtering algorithm to evaluate the test set, and evaluate by RMSE and MAE, as shown in Figure 1.

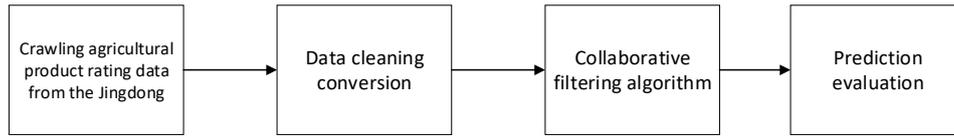


Fig.1 Predicting process

## 2.2 Agricultural Products Rating Data

Using Python to capture the rating data of agricultural products in Jingdong fresh shopping mall, the results are shown in Figure 2.

product_id	publish_time	score	author_name
11434690280	1500114342	4	je****
11434690280	1500129025	5	je****
11434690280	1500127000	5	je****
11434690280	1500127000	5	je****
11434690280	1500227273	1	je****
11434690280	1500226941	4	je****
11434690280	1500242479	5	je****
11434690280	1500282232	4	je****
11434690280	1500282232	5	je****
11434690280	1500303096	5	je****
11434690280	1500307192	5	je****
11434690280	1500312043	5	je****
11434690280	1500313510	5	je****
11434690280	1500316131	5	je****
11434690280	1500316786	5	je****
11434690280	1500323212	5	je****
11434690280	1500329014	5	je****
11434690280	1500329937	5	je****
11434690280	1500337528	5	je****
11434690280	1500339678	5	je****
11434690280	1500416184	5	je****
11434690280	1500417018	5	je****
11434690280	1500419202	5	je****
11434690280	1500417351	5	je****
11434690280	1500494271	5	je****
11434690280	1500499058	5	je****
11434690280	1500505041	5	je****
11434690280	1500577046	1	je****
11434690280	1500607000	1	je****
11434690280	1500677384	1	je****

Fig.2 An example of data acquisition

By sorting and cleaning data, it can be converted into structured data tables, such as (user, item, rating). Each element needs to be represented by a three tuple, namely user, item (agricultural product) and rating. For example, the rating of 46 of agricultural products by user 1 is 1. As shown in Figure 3.

1	46	1
1	47	4
1	48	5
1	49	4
1	50	4
1	53	4
1	54	4
1	55	3
1	56	4
1	61	1
1	62	4
1	65	2
1	66	4
1	68	2
1	69	3
1	89	3
1	96	3
2	1	5
2	3	5
2	4	5
2	5	5
2	6	5
2	7	5
2	8	5
2	9	5

Fig. 3 An example of data exchange

## 3. Collaborative Filtering Algorithm to Predict The Rating

### 3.1 User -based Collaborative Filtering Algorithm

When UserCF(User -based collaborative filtering) predicts the user's rating of an item, it needs to score the item according to the user similar to the user's interest:

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v \in S(u,K) \cap N(i)} w_{uv} (r_{vi} - \bar{r}_v)}{\sum_{v \in S(u,K) \cap N(i)} |w_{uv}|} \quad (1)$$

$\bar{r}_u$  is the mean value of user  $u$  in the training group.  $S(u, K)$  is a collection of  $K$  users most similar to the  $u$  user's interest.  $N(i)$  is a collection of users who are overly evaluated for agricultural  $i$ ,  $r_{vi}$  is the  $v$  user's rating for agricultural  $i$ , and  $\bar{r}_v$  is the average value of all agricultural products that the user has overrated for the agricultural products. The similarity  $w_{uv}$  between users can be calculated by cosine similarity (formula 2) or Pearson correlation (formula 3).

$$w_{uv} = \frac{\sum_{i \in I} r_{ui} \cdot r_{vi}}{\sqrt{\sum_{i \in I} r_{ui}^2} \sqrt{\sum_{i \in I} r_{vi}^2}} \quad (2)$$

$$W_{uv} = \frac{\sum_{i \in I} (r_{ui} - \bar{r}_u) \cdot (r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{ui} - \bar{r}_u)^2 \sum_{i \in I} (r_{vi} - \bar{r}_v)^2}} \quad (3)$$

### 3.2 Item-based Collaborative Filtering Algorithm

ItemCF (Item -based collaborative filtering) predicts that when user  $u$  scores the  $i$  agricultural product, it is required to grade other agricultural products similar to that of the agricultural product  $i$  according to user  $u$ :

$$\hat{r}_{ui} = \bar{r}_i + \frac{\sum_{j \in S(i,K) \cap N(u)} w_{ij} (r_{uj} - \bar{r}_j)}{\sum_{j \in S(i,K) \cap N(u)} |w_{ij}|} \quad (4)$$

$\bar{r}_i$  is the mean value of item  $i$  being overrated by users in the training set.  $S(i, K)$  is the collection of  $K$  agricultural products which are most similar to agricultural product  $i$ .  $N(u)$  is a collection of agricultural products with excessive user  $u$  evaluation.  $r_{uj}$  is the  $u$  user's rating for agricultural product  $j$ ,  $\bar{r}_i$  is the average of agricultural products, and  $w_{ij}$  is the similarity between agricultural products.

Calculating the similarity of agricultural products also adopts cosine similarity and Pearson correlation:

$$w_{ij} = \frac{\sum_{u \in U} r_{ui} \cdot r_{uj}}{\sqrt{\sum_{u \in U} r_{ui}^2 \sum_{u \in U} r_{uj}^2}} \quad (5)$$

$$w_{ij} = \frac{\sum_{u \in U} (r_{ui} - \bar{r}_i) \cdot (r_{uj} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{ui} - \bar{r}_i)^2 \sum_{u \in U} (r_{uj} - \bar{r}_j)^2}} \quad (6)$$

### 3.3 Slope One Algorithm to Predict the Rating

Slope One algorithm is a special item-based collaborative filtering recommendation algorithm. The Slope one algorithm is only applicable to rating predictions. The traditional collaborative filtering algorithm performs score prediction by calculating the similarity of users or items, while the Slope One algorithm does not calculate the similarity between items but uses a simple linear regression to predict. Two step calculations:

(1) The mean value of the difference between agricultural products is defined as the rating deviation between agricultural products; the average deviation value of the two agricultural products  $i$  and  $j$ , the average deviation value  $R(ij)$  of the  $i$  and  $j$  agricultural products is calculated as follows:

$$R(ij) = \frac{\sum_{u \in N(i) \cap N(j)} (r_{ui} - r_{uj})}{|N(i) \cap N(j)|} \quad (7)$$

where  $r_{ui}$  is the rating of user  $u$  on agricultural product  $i$ , and  $r_{uj}$  is the rating of user  $u$  for agricultural products  $j$ .  $N(i)$  is an excessive user of agricultural product  $i$ , and  $N(i) \cap N(j)$  is the user of both of agricultural products  $i$  and  $j$ .  $|N(i) \cap N(j)|$  is the number of users who have overrated agricultural products  $i$  and  $j$ .

(2) predict user ratings of agricultural products through rating errors and user ratings.

$$p_{uj} = \frac{\sum_{i \in N(u)} |N(i) \cap N(j)| (r_{ui} - R(ij))}{\sum_{i \in N(u)} |N(i) \cap N(j)|} \quad (8)$$

where  $N(u)$  is an agricultural product that user  $u$  rated excessively.

## 4. Evaluation Results

### 4.1 Evaluation of UserCF and ItemCF

Different  $K$  values are taken respectively, cosine similarity and Pearson correlation are used to calculate MAE and RMSE. The results are shown as shown in figures 4.

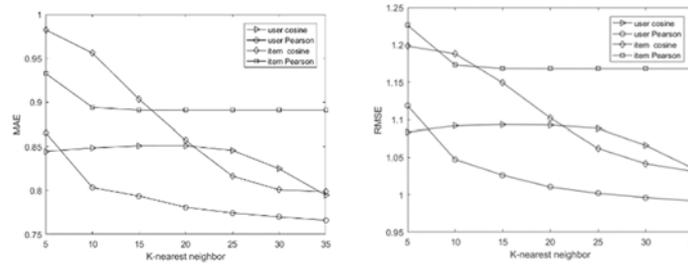


Fig. 4 MAE and RMSE value

As can be seen from figure 4, ItemCF's MAE and RMSE seem to be inferior to UserCF's. However, when choosing which algorithm, the off-line test results do not play the final role. If a reasonable recommendation is required, ItemCF must be chosen. Moreover, if users are numerous, it is difficult to calculate the similarity matrix of users, so it is very difficult to implement, and will not use UserCF. The selection of neighborhood K value has a great influence on the quality of recommendation. As you can see from the diagram, with the increase of the nearest neighbor K value, the range of MAE and RMSE changes is smaller. Therefore, the near neighbor K is generally set to about 25, because the number of agricultural products purchased by the user is not much.

## 4.2 Evaluation of Slope One

The Slope algorithm is simple in computation and is specifically aimed at prediction scores. Compared with other algorithms, the biggest advantage of Slope One algorithm is simple operation, high efficiency, and even can be used for online computing. After testing, the MAE and RMSE values of the Slope One algorithm are 0.9603 and 1.1630 respectively.

## 5. Conclusion

In this paper, the collaborative filtering algorithm is used to predict the data of agricultural products and compare the MAE and RMSE values of various algorithms. The results show that the user based collaborative filtering algorithm is superior to the collaborative filtering algorithm based on items in performance, but in actual application, the appropriate algorithm will be adopted according to the actual requirements. Slope one algorithm is simple in calculation and is specifically for scoring prediction model.

## Acknowledgments

We acknowledge the support of the National Natural Science Foundation of China (No. 61204127), Heilongjiang Education Science "Twelfth Five-Year" Plan for the Record (GBGH27), Science and Technology Project of Qiqihar (GYGG-201619, GYGG-201409), and New Century Excellent Talents in Heilongjiang Provincial University (1253-NECT025). This work was supported by the Supercomputing Center of Qiqihar University.

## References

- [1]. Tianqi ZHANG. Paths for Upgrade and Transformation of E-commerce of China's Fresh Agricultural Products Based on Whole Industry Supply Chain[J]. *Asian Agricultural Research*,2017,9(08):1-4.
- [2]. ZHU Qun-fang, JIANG Chun-sheng. Development mode of the E-commerce of eco-agriculture products in China[J].*Ecological Economy*,2017,13(03):279-290.
- [3]. Hyunwoo Hwangbo, Yang Sok Kim, Kyung Jin Cha. Recommendation system development for fashion retail e-commerce[J]. *Electronic Commerce Research and Applications*,2018,28.

- [4]. Fu M, Qu H, Yi Z, et al. A Novel Deep Learning-Based Collaborative Filtering Model for Recommendation System[J]. IEEE Transactions on Cybernetics, 2018, PP (99):1-13.
- [5]. Ludovico Boratto, Salvatore Carta, Gianni Fenu. Investigating the role of the rating prediction task in granularity-based group recommender systems and big data scenarios[J]. Information Sciences, 2017, 378.