



Big Data-Based Recommendation Algorithm in E-commerce Personalized Marketing

Shujun Li^(✉), Li Li, Yiwen Cui, and Xueyan Wu

School of Economics and Management, Wuhan Railway Vocational College of Technology,
Wuhan 430205, China
sherrylee616@qq.com

Abstract. E-commerce recommendation algorithm is the core of the entire recommendation system, which plays a very important role in e-commerce personalized marketing. Its recommendation accuracy and efficiency directly affect the overall performance of the recommendation system. E-commerce recommendation algorithm based on data mining technology, in-depth analysis of various user data especially user access data, get each user's hobbies, interests and specific buying behavior characteristics. This paper analyzes the related technologies and algorithms of e-commerce recommendation system, and proposes the architecture of e-commerce recommendation system based on user behavior data. In order to meet the requirements of recommendation accuracy and real-time performance, the recommendation module designed in this paper is mainly composed of three modules: content-based recommendation module, collaborative filtering algorithm-based recommendation module and user behavior-based recommendation module, and the functions and technologies of each part are specifically analyzed. Finally, a personalized marketing scenario is created to evaluate the effect of the recommendation system.

Keywords: Data Mining · Personalized Recommendation Algorithm · Collaborative Filtering

1 Introduction

In the process of e-commerce personalized marketing, the recommendation system provides customers with product information and suggestions, helps users decide what products to buy, and simulates sales personnel to help customers complete the purchase process. Personalized recommendation is to recommend the information and goods that users are interested in according to their interest and purchase behavior. The e-commerce recommendation algorithm based on data mining technology deeply analyzes various user data, especially user access data, through data mining technology, and obtains their hobbies, interests and specific purchase behavior characteristics [1]. It generally includes the learning stage and the application stage. In the learning stage, the data mining system analyzes the data and establishes the corresponding recommendation model to explain the user's behavior patterns, also known as the pattern mining stage; In the application

phase, the recommendation algorithm provides users with real-time recommendation services according to the established recommendation model and user behavior, also known as the recommendation generation phase.

2 Common Recommendation Algorithms in E-commerce Personalized Marketing

2.1 Content Based Recommendation Algorithm

It is often used to integrate these behaviors according to the user's behavior history information to calculate the user's preferences, and then recommend the most similar content based on the user's preferences. The advantage of this method is that it does not need to consider other user characteristics, but its disadvantage is that the content available for analysis is limited. The biggest feature is excessive characterization, which may lead to the lack of innovation in the recommended content. Content based recommendation generally includes three processes:

Step 1: Content representation: extract some features for each item to represent the item;

Step 2: Feature learning: using the feature data of an item that a user likes or dislikes in the past to summarize the user's preferences;

Step 3: Generate a recommendation list, and recommend a group of items with the greatest relevance to the user by comparing the user profile obtained in the previous step with the characteristics of candidate items [2]. If we use a classification model in feature learning, we can simply return many items predicted by the model that are most likely to be of interest to the user as recommendations.

Taking the ant colony clustering algorithm as an example, its main architecture is as follows:

- (1) Offline part: mainly use the ant colony clustering algorithm to prepare data, conduct data preprocessing, and get the recommendation pool. First, the user's access records are cleaned and identified offline, and the recommendation pool is obtained by clustering analysis according to the user's access paths to different commodities.
- (2) Online part: mainly use the engine for recommendation. The recommendation system consists of three modules, namely user agent module, recommendation content generation module and recommendation generation module. As shown in Fig. 1.

2.2 Collaborative Filtering Recommendation Algorithm

The recommendation algorithm based on collaborative filtering is mainly used for prediction and recommendation. The algorithm finds user preferences by mining users' historical behavior data, divides users into groups based on different preferences, and recommends products with similar tastes. Collaborative filtering is to compare some behaviors and attributes of some users with those of other users, classify users with high similarity, and then the recommendation system can recommend a product to similar

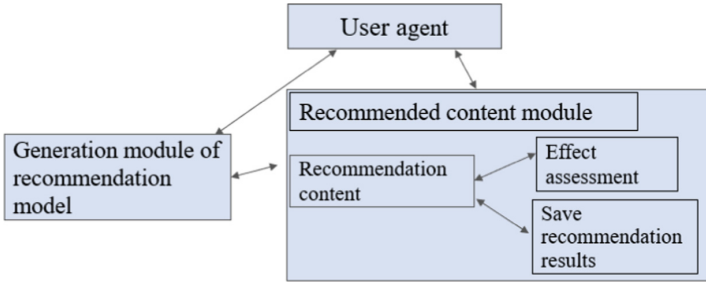


Fig. 1. Online architecture based on ant colony clustering algorithm

users. The accuracy of this recommendation system is significantly improved. However, each product must be purchased by many users before it can be recommended to other nearest neighbors [3]. In this way, some newly added products are difficult to be recommended. This problem is also known as the “cold start” problem of collaborative filtering. The recommended collaborative filtering algorithm is shown in Fig. 2.

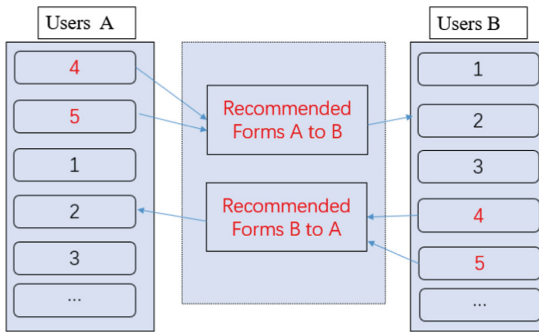


Fig. 2. Collaborative filtering algorithm recommendation

Collaborative filtering recommendation algorithms are divided into two categories, namely user based collaborative filtering algorithm and commodity based collaborative filtering algorithm. It performs well when the scoring matrix information is dense, and can capture some complex information, which often leads to unexpected surprises in practical applications, but it is not suitable for serious recommendation tasks such as public fund recommendation.

2.3 Recommendation Algorithm Based on Association Rules

Association rule-based recommendation algorithm can be divided into offline association rule recommendation model building stage and online association rule recommendation model application stage. In the offline phase, various association rule mining algorithms are used to establish association rule recommendation models; in the online stage, users

are provided with real-time recommendation services according to the established association rule recommendation model and their purchase behavior. The offline establishment of association rule recommendation model can ensure the real-time performance of the algorithm. This algorithm uses data mining technology to obtain rules from a large number of past transaction data, which can be the association rules between goods purchased at the same time, or the sequence model of goods purchased in chronological order [4]. This algorithm is simple in calculation, but it is difficult to recommend commodities without association rules or sequence models.

3 Personalized Recommendation System for E-commerce Products Based on Data Mining

3.1 Overall Framework Design

The database in the server stores a large number of users' web page access path information and search keyword information data, which reflect users' search intention. Our mining and analysis of these data will greatly improve the efficiency of users' search for goods, thus improving users' satisfaction with e-commerce marketing, and thus promoting product sales. According to the analysis of compatibility and other issues, the functions of the recommendation system based on data mining include:

- (1) Have a mechanism to promote new products.
- (2) Analyze customer behavior sequence.
- (3) Intelligently analyze the customer's keyword database and search path.
- (4) Analyze customer behavior sequence. That is, analyze the products that customers may like within the specified time and organize them into a recommendation form.
- (5) A recommended log can be saved regularly.
- (6) Effectively analyze users' characteristics.

The above functional requirements analysis shows that the recommendation system based on data mining needs to avoid affecting the original online engine as much as possible, which requires minimizing the coupling between the recommendation system and the original online engine [5]. The framework of the recommendation system is shown in Fig. 3.

3.2 Design of Recommendation Modules

Content-Based Recommendation Module. Content-based recommendations are based on the similarity between projects. This kind of recommendation first needs to analyze the content of the items that customers have scored, generate a customer information archive, then list the items similar to the new archive from these existing items, sort the selected items (according to the rating), and combine the customer feedback information to recommend. Content based recommendation system is based on the comprehensive consideration of product information, customer information and users' interest in products, so as to obtain a list of recommended products after filtering by the recommendation system and thus personally recommend products to users. Figure 4 shows the content-based recommendation module.

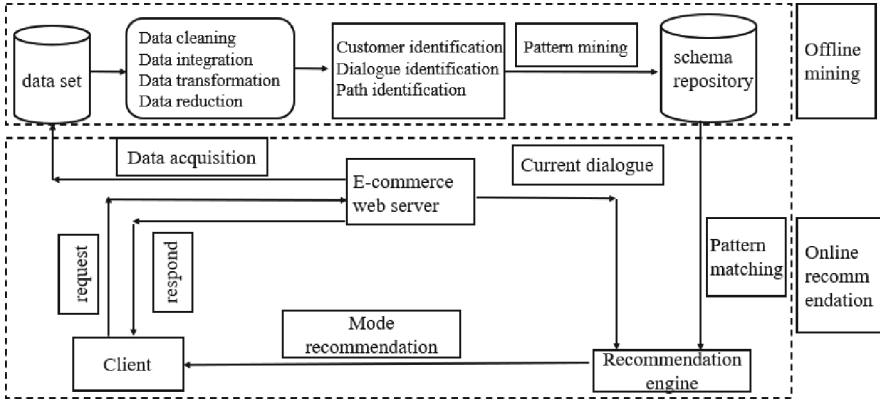


Fig. 3. Architecture of e-commerce personalized recommendation system

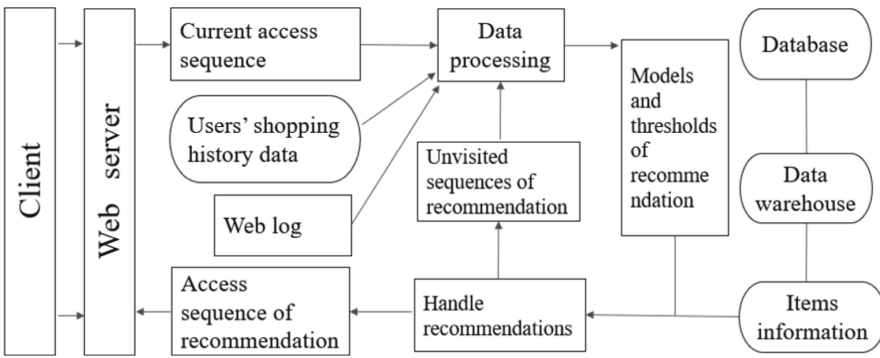


Fig. 4. Content-based recommendation module

Recommendation Module of Collaborative Filtering. (1) Main steps:

- 1) Obtain customer information. This part mainly obtains the customer’s interest rating of the project.
- 2) Analyze the interest similarity between different customers. This part is to analyze the similarity of interests between customers to find the nearest neighbor.
- 3) From the nearest neighbor generated in step (2), find the items that users may like, and recommend them to new users according to their ratings. The structure of collaborative filtering personalized recommendation system is shown in Fig. 5.

The recommendation system needs to select users, find potential buyers, discover their potential purchase value, make them target users, and find similar content that users are interested in, so as to find the nearest neighbor and focus on recommending similar products to these users.

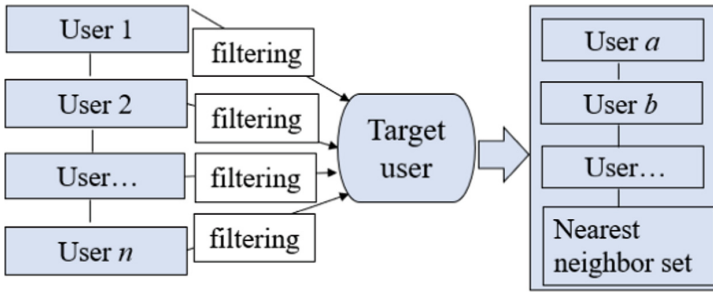


Fig. 5. Structure of collaborative filtering

(2) User-based collaborative filtering recommendation algorithm

User based collaborative filtering recommendation is to generate a Top-N recommendation list for target users based on the interests of neighbor users. It is based on the theoretical assumption that users who like similar items may have the same or similar preferences. User-based collaborative filtering recommendation uses statistical techniques to search for several nearest neighbors of the target user, and then predicts the target user's rating on the unsealed items according to the rating of the nearest neighbor, and selects the first few items with the highest prediction rating as the recommendation results to feed back to the user [6]. The rating of the user-based collaborative filtering recommendation algorithm is to generate the final recommendation result through the nearest neighbor's rating. The current user's rating on the unrated item is approximated by the weighted average of the nearest neighbor's rating on the item. User-based collaborative filtering recommendation algorithm can be divided into the following three stages:

The first stage: data representation. The user rating data can be represented by an $m * n$ -order matrix $R(m, n)$. Row m represents m users, column n represents n items, and element R in row i and column j represents the user i 's rating on item j .

The second stage: nearest neighbor query. The nearest neighbor query is to search for several users whose scoring behavior is similar to that of the current user. The similarity between user a and user b is recorded as $\text{sim}(a, b)$. The main methods to measure the similarity between users are:

- 1) Cosine similarity: the user's rating is regarded as a vector in the n -dimensional item space. If the user does not score the item, the user's rating for the item sets to 0. The similarity between users is measured by the cosine angle between vectors. Let the ratings of user a and user b on the n -dimensional term space be expressed as vectors \vec{a} and \vec{b} respectively, then the similarity $\text{sim}(a, b)$ between user a and user b is as follows:

$$\text{sim}(a, b) = \cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|}$$

- 2) Adjusted Cosine Similarity: This method considers the rating scale of different users. If the set of items scored by user a and user b as I_{ab} , I_a and I_b represents the set of items scored by user a and user b respectively, so the similarity between user a and user b is:

$$sim(a, b) = \frac{\sum_{c \in I_{ab}} (R_{ac} - \bar{R}_a)(R_{bc} - \bar{R}_b)}{\sqrt{\sum_{c \in I_a} (R_{ac} - \bar{R}_a)^2} \sqrt{\sum_{c \in I_b} (R_{bc} - \bar{R}_b)^2}}$$

R_{ac} —user a 's rating of item c

R_{bc} —user b 's rating of item c

\bar{R}_a —Average rating of user a

\bar{R}_b —average rating of user b

- 3) Correlation: the set of items scored by user a and user b as I_{ab} , then the similarity between user a and user b measured by Pearson correlation coefficient method is:

$$sim(a, b) = corr_{ab} = \frac{\sum_{c \in I_{ab}} (R_{ac} - \bar{R}_a)(R_{bc} - \bar{R}_b)}{\sqrt{\sum_{c \in I_{ab}} (R_{ac} - \bar{R}_a)^2} \sqrt{\sum_{c \in I_{ab}} (R_{bc} - \bar{R}_b)^2}}$$

R_{ac} —user a 's rating of item c

R_{bc} —user b 's rating of item c

\bar{R}_a —Average rating of user a

\bar{R}_b —average rating of user b

The goal of nearest neighbor query is to find the user set $C = \{C_1, C_2 \dots C_K\}$, in the entire user space for each user u , so that $u \notin C$, and the similarity $sim(u, C_1)$ between C_1 and u is the highest, the similarity $sim(u, C_2)$ to C_2 and u takes the second place, and so on.

The third stage: recommendation generation. According to the rating information of the current user's nearest neighbor on the item, the current user's rating on the unrated item is predicted to generate the top-N item recommendation [7].

Set the nearest neighbor set of user u as NN_u , then user u will give item i a prediction rating of $P_{u,i}$, which can be got from NN_u rating of user u . The calculation method is as follows.

$$P_{u,i} = \bar{R}_U + \frac{\sum_{n \in NN_u} (R_{n,i} - \bar{R}_n) * sim(u, n)}{\sum_{n \in NN_u} (|sim(u, n)|)}$$

The system uses the above method to predict the user's rating on all the items that have not been scored, and then selects the top N items with the highest predicted rating as the recommendation results to feed back to the current user.

Recommendation Module Based on User Behavior. The recommendation module based on user behavior is mainly divided into three parts in terms of architecture:

- (1) Offline sorting.
- (2) Data and index section.
- (3) The online collection and analysis section.

The relevant functions of each part are described as Fig. 6.

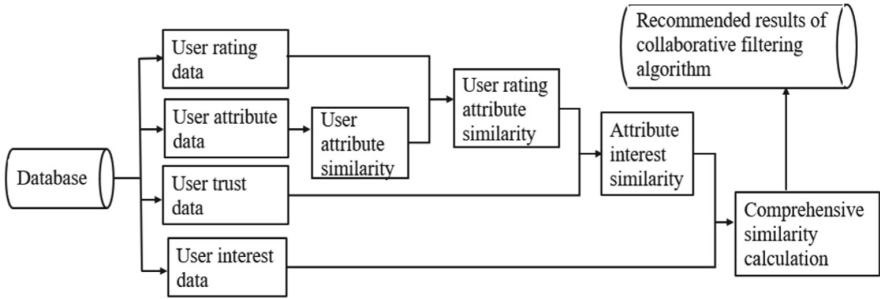


Fig. 6. Architecture of user behavior data-based recommendation

4 Effect and Evaluation

In order to verify the recommendation effect of this system, a small clothing sales website is selected as the experimental environment to observe the effect of different number of users. By comparing the coverage of content-based, collaborative filtering, user behavior-based recommendation and the recommendation system designed in this paper based on data mining, the average satisfaction of the four recommendations is evaluated.

In the experiment, when the system recommends, if the users put it in favorites or place an order, it is deemed that users are satisfied with the recommendation results; if the user no longer needs to recommend after selecting, it is deemed that he is dissatisfied with the result; if the user clicks in to view the recommended products, but does not place an order, it is considered as normal [8]. The experiment is divided into two parts. The 1st part tests the accuracy of the system, that is, the user's satisfaction with the products recommended by the system. The 2nd part is the recommendation rate, that is, to test the coverage of products recommended by the system. The 2nd part checks the coverage of the three recommendation algorithms (including the proportion of users' ideal commodities) when 90, 120 and 200 items are taken out, and the results are shown in Table 1:

Table 1. The coverage of four recommended algorithms

	content-based	behavior-based	collaborative filtering	integration algorithm
90 items	77.4%	86.3%	81%	90.1%
120 items	75%	82%	77.8%	90.2%
200	72.6%	83.4%	75%	89.9%

The results of the 1st part show that with the increase of the number of users, the satisfaction of each recommendation is higher and higher. In the case of the same number of users, the behavior-based and content-based recommendation is similar, and the satisfaction of collaborative filtering recommendation is higher than that of the two recommendations [9]. However, after the integration of the three recommendations, the satisfaction is significantly improved. The 2nd part of the experiment shows that with the increase of the number of goods, the recommendation coverage of all strategies will decrease, but for the same number of items, the coverage of the integration algorithm is significantly higher than that of other strategies. To sum up, the system based on data mining and integration algorithm has significantly improved the recommendation rate and accuracy.

5 Conclusion

In e-commerce activities, compared with traditional marketing methods, personalized marketing is more targeted because it can carry out precise marketing according to the individual differences. Personalized recommendation technology is the key of e-commerce recommendation system. The recommendation technology based on data mining is based on users' habits, hobbies and interests, which can more easily recommend appropriate products, and thus improve the number of visits, clicks and orders on the website. The personalized recommendation system based on data mining explored and designed in this paper can not only effectively solve the problem of huge and messy information in the recommendation system, but also realize the personalized presentation of items, which has great application research value. However, there are still many aspects to be improved. For example, the flexibility of the recommendation system needs to be further verified, the adjustment of rating types and the corresponding calculation need to be further improved, and the fusion of the recommendation lists generated by the three algorithms is relatively stiff, which needs further improvement.

References

1. Mobasher, B., Cooley, R., Srivastava, J.: Automatic personalization based on Web usage mining. *Commun. ACM* **43**(8), 142–143 (2000)
2. Yang, F.: Research on E-Commerce Personalized Recommendation Technology Based on Data Mining. Xi'an University of Electronic Science and Technology, Xi'an (2008). (in Chinese)
3. Pei, L.: Research on E-Commerce Recommendation System Based on Web Data Mining. Tongji University, Shanghai (2006)
4. Qu, T.: Research and Implementation of Web Data Mining Technology in E-Commerce System. University of Electronic Science and Technology, Chengdu (2011). (in Chinese)
5. Wang, W.: E-Commerce Personalized Recommendation Technology Based on Data Mining. University of Electronic Science and Technology, Chengdu (2014)
6. Guo, X.: Big data precision marketing based on recommended algorithms. *Inf. Technol. Stand.* **05**, 40–41 (2019). (in Chinese)
7. Liu, F.: Design and Implementation of E-Commerce Personalized Recommendation Algorithm. Jiangsu University, Nanjing (2010). (in Chinese)

8. Li, J.: Application of e-commerce personalized information automatic recommendation algorithm based on big data technology. *Autom. Technol. Appl.* **10**, 38–39 (2021). (in Chinese)
9. Nascimento, G., Correa, R.F.: Evaluation of selection criteria for noun phrases with relevance for information retrieval. *Transnormal* **30**(2), 179–184 (2018)