Precise Recommendation System for Agricultural Products in E-Commerce

Chao Xie
School of Management, Wuhan Donghu University, Wuhan 430212, Hubei, China

Qiongjie Zhou*
School of Management, Wuhan Donghu University, Wuhan 430212, Hubei, China
*Corresponding author (Email: zhouqj77katharine@163.com)

Defu He
College of Computer, Hubei University of Education, Wuhan 430205, Hubei, China

Abstract
To solve the problem of low precision of Individualized Recommendation of agricultural products in e-commerce, the Precise Recommendation System for Agricultural Products in E-commerce is researched. Aiming at the insufficiency of the classic Apriori algorithm, a new weighted fuzzy association rules mining algorithm is put forward to ensure the downward closure of frequent itemsets. The workflow of the recommendation system was tested through the structural design of e-commerce recommendation system, data preprocessing module design and recommendation module design. The hit rate is selected as the evaluation standard of different recommendation models. The contrastive analysis for the practical collected data was conducted with the half-off cross test method. The experimental results show that the hit rate of the Top-N products in the association rule set is significantly higher than that of the interest recommendation method and the best-selling recommendation method.

Key words: Weighted Association Rule, Mining Algorithm, Electronic Commerce, Recommendation System

1. Introduction
The popularization of computer and network technology provides convenient conditions for the development of e-commerce. With the increasing number of netizens and online trading activities, e-commerce has become one of the most important emerging industries in China. At present, the development of rural e-commerce is also an important form of a combination of e-commerce and traditional agriculture, which has an important positive impact on rural economic construction and the improvement of farmers’ income level. In 2017, rural online retail sales reached 124.48 billion RMB, an increase of 39.1% over the previous year. It was expected that the rural online retail sales would exceed 1.6 trillion RMB in 2018, with an increase of more than 35%. In addition, in 2017, the net retail sales of 832 poverty-stricken counties reached 120.79 billion yuan in 2017, which is an increase of 52.1% over the previous year was 13.0% higher than the rural growth rate. It was precisely in view of the tremendous changes which brought by e-commerce to some rural areas that the State Council Poverty Alleviation Office formally incorporated “e-commerce poverty alleviation” into the policy system of poverty alleviation in 2014 and implemented it as one of the “ten projects of precise poverty alleviation” in 2015. With the rapid development of rural e-commerce industry, the number of users purchasing agricultural products on the Internet has exceeded 400 million. In the huge sales of e-commerce agricultural products, shoppers and service providers are facing one problem that users and businessmen cannot clearly identify the required agricultural products; mass of agricultural products cannot quickly and accurately match customer preferences; screening commodities last for too long. These problems have greatly hindered the development of the e-commerce industry. In order to solve these problems, this paper makes use of relevant knowledge in mathematics to study the precision recommendation system of e-commerce agricultural products based on weighted association rule mining algorithm.

2. E-Commerce Recommendation System
With the continuous popularity of the Internet, people pay more and more attention to the convenience, speed, low cost and no limitation by time and space of the e-commerce. However, with the rapid growth of information today, e-commerce websites need to provide users a “procurement assistant”, which can help
e-commerce websites to provide users information and suggestions that meet their needs according to their interests, and help users spend the least time to find the most satisfactory goods and services. The “Purchasing assistant” is the personalized recommendation system of e-commerce.

In 1997, Resnick and Varian have given the formal definition of E-Commerce Recommendation System in Recommender Systems: “Provide information and suggestions to customers on e-commerce websites, simulate sales staff to help users decide what to buy, and help customers successfully complete the purchase process”. Sufficiently recommend suitable products to users according to their historical data, so it is also known as e-commerce personalized recommendation system. This definition has now been widely cited.

The selection of recommendation technology can determine the recommendation performance of recommendation system. It is the core part of the personalized recommendation system in e-commerce. Generally speaking, personalized recommendation technology can be divided into two types: active and passive. Active refers to the analysis of user transaction information and browsing information based on the user's historical data, and then produces recommendation results that satisfy user's preferences. Passive refers to the user querying commodities and related information in the system through certain channels, such as keywords. Inputs, browsing the page, etc. Thus, compared with the active recommendation, the passive recommendation method cannot find the potential interest and purchase intention of users, and the degree of intelligence is low. At present, scholars at home and abroad have done much research on e-commerce recommendation technology and methods. The main research directions and contents focus on the following aspects:

(1) Real-time research: In the recommendation system of an e-commerce website, it is more and more difficult to guarantee the time and quality of recommendation at the same time. Some recommendation technologies improve the timeliness of recommendation while reducing the quality of recommendation. More and more scholars pay attention to how to meet the real-time requirements of recommendation system and ensure the quality of recommendation. The literature solves the real-time problem of recommendation system based on collaborative filtering algorithm [1].

(2) Recommendation quality research: In e-commerce systems, especially in those large-scale systems, customer evaluation data is extremely sparse, which makes the recommendation system unable to produce valuable recommendation results based on these data. To solve this problem, a collaborative filtering algorithm based on rough set and fuzzy clustering is proposed in the literature [2].

(3) Research on the integration of multi-technology and multi-data: At present, most e-commerce recommendation systems only use a small part of data information in their databases to produce recommendation results. However, with the continuous development of e-commerce recommendation system, in order to provide users more effective recommendation services and results, the recommendation system is also moving towards the direction of multi-technology and multi-data integration. Documents use clustering, association and improved collaborative filtering and other recommendation technologies to analyze and process multi-type data. Finally, different recommendation results are given [3].

(4) The wide application of data mining technology: With the deepening of recommendation technology research, recommendation system based on data mining technology has been widely used. Literature improves the mining algorithm from recommendation generation and Web access sequential pattern mining [4].

3. Data Mining and Association Rules

3.1. Data Mining

Data mining, in a broad sense, is a data processing process that has potential, effective value and novelty through a certain algorithm and processing mode in massive data. Data mining patterns can be divided into six types: Association pattern, sequence pattern, classification pattern, regression pattern, time series pattern and clustering pattern. Because of its wide range of applications and strong practicability, association pattern has been greatly developed in the field of modern electronic commerce, and it has become the most important research field in data mining [5-7].

After years of research and analysis, the main methods of data mining include concept description, association analysis, class knowledge mining, and predictive knowledge mining and so on. Although there are many methods of data mining, there are several difficulties in the thorough application of data mining to e-commerce [8, 9]: diversity of data types, poor efficiency and scalability of algorithms, poor interaction of data mining systems, poor data security and privacy. These difficulties hinder the application of data mining in real business.

3.2. Classical Association Rule Mining Algorithms

There are many kinds of association rules mining algorithms in practical application, the most classical of which is the Apriori algorithm. The algorithm has the characteristics of a single dimension, single layer and a
Boolean type. The steps of generating frequent itemsets of the algorithm are as follows:

Step 1: Calculate the frequency of occurrence of all item sets with one element to determine the maximum one-dimensional itemset.

Step k: It is divided into two stages. Firstly, the candidate itemset $C_k$ is generated by the maximum itemset $L_{k-1}$ generated in step $(k-1)$. Then, the support degree of $C_k$ is calculated by searching the database, and $L_k$ is generated according to the minimum support threshold.

The last step is to cycle back and forth in step $k$ until the candidate itemsets cannot be generated, and finally the algorithm ends. The flow chart of the Apriori algorithm is shown in Figure 1.

The main problems of this algorithm for the current e-commerce application are: too many times to scan database; the operation time increases with the increase of the length of frequent items; too large $C_k$ in each stage; unable to update and directly process numerical data; and the mining of association rules in database cannot be directly applied [10-13]. These problems have led to many problems in the application of the Apriori algorithm in e-commerce. For this reason, this paper proposes an improvement of the mining algorithm based on the weighted fuzzy association rules.

4. Improvement of Mining Algorithms Based on Weighted Fuzzy Association Norms

4.1. Weighted Fuzzy Relevance Norm Model

(1) Boolean Weighted Association Rule Model
The Boolean weighted association rule model consists of the transaction set $T$, itemset $I$ and item weight set $W$. When calculating, it can be weighted according to the element attributes and the support degree of Boolean attribute itemset.

(2) Weighted Fuzzy Association Model Rules
The weighted fuzzy association model consists of the transaction set $T$, itemset $I$ and fuzzy set $L$. The membership function of the fuzzy set and the range is $(0, 1)$. The weighted fuzzy confidence of weighted fuzzy
association rule $X \rightarrow Y$ can be calculated according to the ratio of weighted fuzzy support and weighted FITW of fuzzy itemset $X$. The specific calculation process is as follows.

(3) Downward closure of frequent itemsets

Downward closure is the characteristic of the Apriori algorithm, which can generate the maximum frequent itemsets through $K$-item sets. However, in the mining of weighted association rules, because items are given weight attribute and support attribute, the subset of frequent itemsets can no longer determine whether they are frequent.

4.2. Weighted Fuzzy Association Norm Mining Algorithms (NFWARM)

The basic idea of the weighted fuzzy association norm mining algorithm NFWARM is as follows: the algorithm scans the database to get the $C$ value and assigns the weighted fuzzy support attribute to it; the algorithm calculates the candidate item set through a specific function, deletes and prunes the infrequent subset; the algorithm calculates the support degree of candidate item set iteratively until all candidate sets are empty [14-16]. In this way, the frequent itemset is obtained.

Weighted Fuzzy Association Norm Mining (NFWARM) still uses the method of searching the iteration layer by layer to calculate the frequent itemsets, which is the same as the Apriori algorithm [17]. It can be proved by relevant theorems that the algorithm can achieve the downward closure of the frequent itemsets when dealing with weighting.

4.3. Test and Result Analysis

In order to test the advance of the weighted fuzzy association specification mining algorithm NFWARM, this paper carries out relevant data tests. Firstly, ten thousand transactions and attributes are generated randomly through IBM data generator [18, 19]. Among them, there are 20 items in each transaction, each attribute is assigned a weight in $(0, 1)$, and each numerical attribute is reassigned five fuzzy sets. Then the original database is transformed into a fuzzy database by a membership function. The BWARM curve and NFWARM curve represent the execution results of the traditional Boolean weighted association rule algorithm and the improved algorithm respectively. The specific changes are shown in Figure 2.

![Figure 2. Comparison of two algorithms](image)

(a) Comparison of Frequent Item Number with Support Degree

(b) Comparison of Rule Number with Confidence

As can be seen from Figure 2, with the increase of minimum weighted fuzzy support, the number of
frequent itemsets and frequent rules obtained by NFWARM algorithm decreases, while the number of frequent itemsets generated by BWARM algorithm is much less than that of NFWARM algorithm. These data changes show the superiority of the improved NFWARM algorithm.

5. Design and Implementation of E-Commerce Commodity Recommendation System

5.1. Structure Design of Recommendation System

(1) System Development Tools and Platform
The development tools of this e-commerce commodity recommendation system are MyEclipse 9.0, JDK 1.7, Windows 10, and Tomcat 7.0 for the application server and SQLSever 2008 for background database.

(2) Background of E-commerce Platform
The precision recommendation platform for e-commerce agricultural products is mainly an electronic shop which mainly sells agricultural products. With the promotion of rural e-commerce business, the sales volume of the store has been greatly improved. However, with the increase of sales, there are many kinds of agricultural products, and it is difficult for the customer groups to match the suitable agricultural products in a short time. Therefore, the sales volume of the store has met a bottleneck, and it is urgent to develop a high-quality e-commerce precision recommendation system for agricultural products.

(3) Structural design
Precision recommendation system for agricultural products in e-commerce needs to be implemented through two important modules, namely the data acquisition system and data pre-processing system. Data acquisition system mainly provides agricultural products information, while data preprocessing system is responsible for processing customer transaction data and converting it into a data format required by association rules mining algorithm. The core part of the whole system is to use a weighted association rules mining algorithm to build a platform between customer transaction history data and customer purchase demand. The workflow of the whole E-commerce precision recommendation platform for agricultural products is shown in Figure 3.

Figure 3. Structure design of recommendation system

5.2. Data Preprocessing Module Design

(1) Data preparation. The original data such as shop website, agricultural product information, commodity data and transaction information are input into Excel file as the original database. Among them, customer purchase history information and commodity information are the core database of the whole system.

(2) Data classification. The original data stored in Excel files are complex and changeable. It is necessary to classify these original data in order to facilitate data mining. Its classification method is: to abstract the commodity information upward as the mining object of the first conceptual level data; to continue abstracting the commodity information upward again, this data is only represented by symbols.

(3) Data preprocessing. The concept level is used to classify the goods, and the information of goods on and off shelves is processed in time by checking the integrity and consistency of the data.

(4) Generation of historical transaction data. Transaction data is written in SQL scripting language, and the preprocessed historical transaction data is transformed into transaction data mined by joint rule algorithm. At this time, transaction time should be regarded as the variable condition of customer data mining.

5.3. Recommendation Module Design

In the design of recommendation module, customers are required to register as members by landing on the
precision recommendation platform of e-commerce agricultural products, and then the recommendation module is divided into the following two situations according to whether customers have purchased commodities or not.

(1) No customer purchase record

In view of this situation, the most popular Top-N sales strategy is recommended for module design. The design process is as follows: membership registration recommendation of best-selling goods willingness to buy independently rejection, continue to recommend the next level of best-selling goods choose independent goods purchase.

(2) Records of purchases by customers

In view of this situation, it is necessary for the system to collect the purchase records on the platform based on the member information to quickly and accurately calculate the goods that the customer may like to buy and recommend them. The design process is as follows: Member login→Automatic access to purchase history information→Analysis of Customer Hobbies Based on Association Rules Set→Recommend Matched Goods→purchase.

5.4. System Implementation and Testing

(1) On the basis of user’s history transaction records. Customers can search the purchase information by clicking on the “purchase history” after they log into the E-commerce precision recommendation system for agricultural products, and the system background will automatically evolve the rule set. Then, according to the rule set, they can match the garment information in the system and recommend it to customers first.

(2) On the basis of user’s shopping basket information. After customers log into the E-commerce precision recommendation system for agricultural products, according to the shopping basket information generated in the browsing information, the system background will automatically evolve the rule set, and then match the agricultural products information in the system according to the rule set, and recommend it to customers first.

(3) On the basis of user’s no purchase record. For customers without purchase records, the system will recommend the best-selling agricultural products to customers according to the Top-N sales strategy.

6. Recommendation System Result Analysis

6.1. Test Data

In order to ensure the accuracy of the test, the transaction records of agricultural products in five months in 2018 were collected. It includes the basic information of more than 500 commodities and customer information of more than 16,000 Taobao IDs. The commodities involved include rice, tea, fruit and local products.

At the same time, users' buying habits were analyzed before the experiment. Among them, the number of users who only buy one commodity is more than 9,300; more than 5,600 people buy two or more different brands, and more than 1,500 people have the same brand.

After the analysis of the data preprocessing module, the volume of transactions conforming to association rules mining is 5,985.

6.2. Test Evaluation Criteria and Scheme

The evaluation criterion chosen in this experiment is the hit rate. This index can more accurately reflect whether the products recommended by the system will be purchased by customers.

The test scheme is tested by the five-fold cross method. The specific scheme is as follows:

The data is divided into two parts: calculation data and test data. The calculation data is automatically calculated by the algorithm, and the test data is recorded and counted by the customer's real purchase. The data is divided into five parts, four of which are training sets and one of which is testing sets. After each test, five pieces of data are re-combined and tested for five times in the next test.

6.3. Test Results and Analysis

According to the test results, the purchase information of the first 5,000 users is analyzed and the results of five data intervals are obtained. This experiment analyzed the hit rate of Top-N products based on association rule set, Top-N products based on user interest and Top-N products recommended as best-selling products. The results of the comparison are shown in Figure 4.
The average hit rates of the three recommendation strategies are 0.319 for Top-N product based on user interest, and 0.552 for Top-N product based on the association rule set. From the comparison of these three data, we can see that the E-commerce recommendation model based on weighted association rule mining algorithm is effective, and its hit rate is significantly higher than the other two.

7. Conclusions

This paper outlines the basic information of data mining and association rules, points out the shortcomings and shortcomings of the classical Apriori algorithm, and proposes a new weighted fuzzy association mining model algorithm to ensure the downward closure of frequent itemsets. Through the structured design, data preprocessing module design and recommendation module design of the precision recommendation system for e-commerce agricultural products, the recommendation system is completed. Finally, the hit rate is selected as the evaluation criterion of different recommendation models, and the actual data collected are compared and analyzed by the five-fold cross-test method. The test results show that the hit rate of Top-N product in the association rule set is significantly higher than that of interest recommendation and best-selling recommendation, which fully demonstrates the superiority of the recommendation system based on the weighted association rule mining algorithm.

Acknowledgments

This work was supported by the grants from Youth Foudation WuHan Donghu University in 2019 “Research on Traceability System Construction of Quality and Safety of E-commerce Agricultural Products from the Perspective of Blockchain” and Hubei Provincial Collaborative Innovation Centre of Agricultural E-Commerce (Wuhan Donghu University research [2019] No.10 Document).

References


