Research on intelligent recommendation algorithm of e-commerce based on association rules

Jiajie Shen^{1,2,a)} and Xianyi Cheng^{1,2,b)}

^{1}School of Computer Science and technology, Nantong University, JiangSu Nantong 226019, China

^{2} Nantong Research Institute for Advanced Communication Technologies, JiangSu Nantong 226019, China.

E-mail: a)598864151@qq.com, b)Corresponding author: xycheng@ntu.com

Abstract. As the commodities of e-commerce are more and more rich, more and more consumers are willing to choose online shopping, because of these rich varieties of commodity information, customers will often appear aesthetic fatigue. Therefore, we need a recommendation algorithm according to the recent behavior of customers including browsing and consuming to predicate and intelligently recommend goods which the customers need, thus to improve the satisfaction of customers and to increase the profit of e-commerce. This paper first discusses recommendation algorithm, then improves Apriori. Finally, using R language realizes a recommendation algorithm of commodities. The result shows that this algorithm provides a certain decision-making role for customers to buy commodities.

1. Introduction
With the rapid development of Internet and electronic commerce, more and more people are inclined to online shopping. But customers often lose in a large amount of information and become aesthetic fatigue, which lead to not find the goods they need\textsuperscript{[1]}. If businesses can recommend commodities which meet users' demand from mass of commodities to the users, it is expected to transform the potential demand into actual demand. So according to users' characteristics the recommendation system come into being\textsuperscript{[2]}.

The recommendation system has important application value, but it is also faced with many difficult problems in theory and application, among which sparse data, over fitting, extendibility and multimedia information feature extraction are the main problems\textsuperscript{[3]}. And now, although R language can use Apriori algorithm to explore mining, but it can only find the association rules between goods, and cannot achieve the recommended effect. Therefore, this paper presents an intelligent e-commerce recommendation algorithm based on association rules. Through the recent customer behavior analysis, model by Apriori algorithm and mine the relationship between the commodities, and then according to the recent behavior of customers, intelligently recommend products for customers, to play an auxiliary decision-making effect on helping customers to buy commodities\textsuperscript{[4]}.

2. Association rules and improvement of apriori algorithm
The association rules analysis is one of the most active research methods in data mining, the purpose is to find the association relationships between items in a data set that is not directly represented in the data\textsuperscript{[5,6]}. Some definitions of association rules: The support of association rule is the ratio of the number of transactions containing X and Y to all transactions at the same time:
The confidence of association rules is the ratio of the number of transactions containing X and Y to the number of transactions containing X:

\[ \text{conf}(X \Rightarrow Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X)} \]

But the confidence measure ignores the support of the item sets in the rule post, the rules of high confidence may be misleading, so one of the ways to solve this problem is to use the lift degree. The lift degree of association rules represents the ratio of the probability that contains Y after contains X to the probability that contains Y under the condition without X:

\[ \text{lift}(X \Rightarrow Y) = \frac{\text{conf}(X \Rightarrow Y)}{\text{support}(Y)} \]

If the value = 1, indicating that two conditions without any correlation, if the value < 1, X and Y are mutually exclusive, generally in data mining when the value is greater than 3, the rule is valuable.

If the support of an item is greater than the min_sup specified by the user, it is called frequent. The task of association rule mining is to find the rule \( A \Rightarrow B \) which the support and confidence in transaction database D satisfy the min_sup and the min_con specified by the users, that is strong association rules. Association rules mining is a two step process: (1) finding all frequent item sets; (2) generating strong association rules by frequent item sets\(^7\).

Apriori algorithm is a classic frequent item sets mining algorithm. Discovery process of Apriori algorithm is divided into two steps: the first step by iteration, retrieve all the frequent item sets in a transaction database, which supports not less than the threshold value set by users; the second step is the use of frequent item sets to construct rules meet the minimum confidence set by users.

3. Intelligent recommendation algorithm for goods

The implementation steps of improved intelligent e-commerce recommendation algorithm based on association rules as shown in Fig.1, the original customer transaction records as the data source for the recommendation system, through the data pretreatment, will be converted into standard customer transaction data. Based on a large number of user transaction data, association rule model for the above research as the basic algorithm, looks for user transactions including the purchase mode and buying interest, in the form of association rules stored in the rule base. After obtaining the user inquiry or purchase information, according to the association rule, the recommendation is produced, and the commodity which may be of interest to the users is recommended to the users in real time.

![Figure 1. Flow chart of e-commerce intelligent recommendation algorithm based on association rules](image)

3.1 Data pretreatment

According to three months of customer information provided by an electricity supplier website, the name of each field is explained in Table 1.
Table 1. Description of customer behavior information

| user_id | User number | desensitization |
|---------|-------------|-----------------|
| sku_id  | Commodity number | desensitization |
| time    | Behavior time   |
| type    | 1. browse (refer to browse details page); 2 plus; 3 shopping cart; 4 deletion; 5 orders; 6 Click; |

Because the original data has a large number of missing, inconsistency and abnormal data, if the data is not removed, it will lead to deviation of mining results, so data pretreatment is very important.

By comparing the two dimensions, it is very clear that there is 3672710 duplicate data in the original data. If it is not removed before mining, it will seriously affect the result of the mining.

3.2 Implementation of recommendation algorithm based on apriori

Then the transaction record of each user is constructed. First, removing the other columns, only leaving the user ID and the commodity ID, because only two columns are needed to build the transaction data. Here's the transaction data for type=1 and type=4, because it's better to understand what you'll buy after you've browsed that product, or what you'll browse after you buy it. Then use Apriori algorithm to mine the transaction database constructed to find association rules, create a set of association rules (setting the minimum support and minimum confidence, and in accordance with the lift value from high to low ranking). Association rules are shown in Fig.2. Because general default rule which lift value is greater than 3 is valuable, select the top 4 rules in the rule base.

After having the association rules base and the customer transaction database, first create a recommendation set for each user and initialize it, the lhs in rules is matched with the items in the transaction data through the loop by the defined function. Once the match is successful, the rhs in the rules will be recommended to the user, which is exactly what the Apriori algorithm lacks in the R.

4. Conclusions

In this paper, the related knowledge of recommendation algorithm based on association rules is discussed. Based on the research of traditional Apriori algorithm, the recommendation algorithm is studied. By selecting the rules of high lift value, we can effectively avoid the misleading of high confidence, and then find the valuable association rules, and then recommend the customer's interested goods to the customer through the recommendation algorithm. The experiment proves that the algorithm is effective and feasible, and it can provide data support for the management decision of the merchant, and it is helpful to improve the service level and quality, and also provides a basis for the
customer to buy the goods.

Acknowledgments
This paper is supported by the National Natural Science Foundation of China (No. 61340037).

References
[1] Yongjie Yang, Pengfei Yang, Qun Duan, Lina Han. Intelligent recommendation algorithm of goods based on association rules [J]. Modern computer, 2016, 10 (04), pp. 25-27.
[2] Xuan Li, Xiaoyan Wang, Zhenquan Zhuang. Personalized intelligent recommendation service based on Association Rule Mining [J]. Computer Science, 2002, 38 (7), pp. 200-204.
[3] Bo Yang, Peng-fei Zhao. Overview of recommendation algorithms [J]. Journal of shanxi university, 2011, 34 (3), pp. 337-350.
[4] Yu-lin Liu. Research on hot spot of combination sale of e-commerce based on association rule [J]. Journal of Zhejiang Sci-Tech University, 2016, 6(36), pp. 528-541.
[5] Xue Ding. Research on Intelligent Recommendation System of books based on association rules [J]. Library Information Forum, 2010, 4, pp. 48-52.
[6] Carlos Ordonez, Norberto Ezquerra, Cesar A.Santana. Constraining and summarizing association rules in medical data [J]. Knowledge and Information Systems, 2006, 3(9), pp. 259-283.
[7] Jan Rauch. Logic of Association Rules [J]. Applied Intelligence, 2005, 22 (1), pp. 9-28.