A Method Combined with User's Scores for Optimizing Association Rules

Qinglie Wu[1,2](✉) and Yu Yang[1]

[1] School of Economics and Management, Southeast University, Nanjing, China
wql@seu.edu.cn

[2] Jiangsu Academy of Smart Industries and Digitalization, Nanjing, China

Abstract. In personalized recommendation services, association rules are often used to provide users with appropriate recommendations. However, it is often difficult to choose the results of the optimal rule. Association rule recommendation methods, usually only need to make use of the transaction data set, and do not use many more accessible domain knowledge, the recommendation results are not satisfactory. This paper combines the collaborative filtering thinking and uses the user score information to propose a new rule result selection method, which combines the rules with consistent recommendation results, and then determines the similarity between the user and the combined rules by using user's Scores, and then obtains the optimal recommendation result. Experimental results show that the proposed method has better user satisfaction compared to traditional methods.

Keywords: association rules · user's Scores · optimization method

1 Introduction

With the rapid development of computer and Internet technology, most businesses will pay attention to online sales, and the scale and influence of e-commerce applications are also expanding. Internet applications are gradually changing people's lifestyle and habits, and the number and types of online goods are growing rapidly. Traditional information retrieval methods cannot help users to accurately find out the goods and services they need, and personalized recommendation methods have emerged in this environment [4, 12, 13]. Be advised that papers in a technically unsuitable form will be returned for retyping. After returned the manuscript must be appropriately modified.

Association rule recommendation is an important kind of personalized recommendation method. Through the mining of the transaction database, it obtains the potentially valuable rule information hidden in the massive data. This method is favored by users because of its less use of user information and can complete the initial calculation offline [1, 9]. However, after using the associated rule recommendation method, it is often difficult to select the optimal rule to recommend to users.

Using the association rule recommendation, a large number of rule results are usually obtained. If the rules are only selected through the value of support and confidence, it

J. Yang et al. (Eds.): ICMSEM 2022, AHE 12, pp. 160–172, 2023.
https://doi.org/10.2991/978-94-6463-038-1_17

is difficult to select the appropriate results. In a shopping website, for example, there is a need to recommend target customers, site recommended area, only allowed to push a product to the user, and the rules result set, there are multiple support and confidence equal rules meet the conditions, the rules recommended results are not the same, then should choose which rule to recommend to the user? Obviously, in some cases, the basic metric of the association rule algorithm cannot yield the best recommendation rule.

At present, some results have been obtained on the study of association rule results. For the problem of too many rules results to be used effectively, some processing methods, such as clustering [2, 6, 8], visualization [10], deleting redundant [7, 11], can alleviate the rule selection problem caused by the large number of rules results to a certain extent.

However, none of these methods effectively utilize domain knowledge, which may make the recommendation results less satisfactory. At the same time, this method does not solve the above problems encountered in the selection results of support-confidence indicators. Jiang et al. (2010) [5] and Liu et al. (2018) [7] proposed a method to solve the difficult problem proposed in this paper, using group decision ideas, when gathering association rules, proposed "identical, inconsistent" as experts, using evidence theory to gather, and apply this method to the field of computer purchase recommendation, from the perspective of the correlation between customer preference and satisfaction, make the rule information is fully utilized. However, the algorithm needs to extract the demand——satisfaction dataset from the customer text comments, which is more difficult to obtain, and needs to involve some customer personal information.

The main work of this paper is to study and propose specific solutions to the recommendation problems of the above association rules. First, the specific idea of association rule optimization is proposed, and then the specific steps of association rule optimization are discussed in detail, and the effectiveness of the proposed association rule optimization method is verified through calculation case analysis and experimental analysis.

2 Optimization Ideas of Related Rules

The biggest advantage of association rule recommendation is that it can achieve better recommendation results without too much user information. In today's situation where privacy information is very important, it is unrealistic to obtain more detailed user information, and users' scores and evaluation information is generally more easy to obtain. Mainstream shopping websites, music and movie websites will have a user five-star standard evaluation information. Obviously, the combination of user evaluation information can improve users' recommendation satisfaction. The idea of the association rule algorithm is to search for the frequent transactions in the database, and then analyze the frequently occurring projects in the combination, and then get the recommended results. It is generally difficult to combine the user's evaluation information into the algorithm subject, but it is feasible to solve the above several problems by combining the user's score information. Group decision-making is a very effective method of choice, which gathers the views and preferences of all the members of a group on something, and then makes a certain order to make the final optimal recommendation [3]. Combining the recommendation results of association rules with users 'scoring and group decision-making

ideas can effectively solve the shortcomings of the existing post-processing methods of association rules' results.

Based on the analysis as described above, we propose a specific idea of optimizing the correlation rules based on the user score: First, we should compare the recommended target user behaviour with the rules in the rule base, and find out the eligible rules. Then put the same recommendation result rules together, and composites a recommendation rule set, so, each rule set has a number of rules forming a small group. Then we calculate the score of the rule precursors within these rule sets, and use the mean of the user's score as the score value of the rule. This way these rules can be viewed as neighbour users of the target users in the collaborative filtering algorithm. To compare the similarity of each rule precursor to the target user's score within the group, and predict the target user's score value on the recommended results. Finally, the recommendation result with the highest score value was used as the optimal choice. This process combines a collaborative filtering algorithm and group decision thinking.

The above optimization ideas can help solve the problem of consistent anterior and posterior rules in the rule results, and can also solve the difficulty of rule selection caused by different support and confidence sizes, and make full use of all the rule information in the rule library.

3 Associative Rule Optimization Steps

According to the above specific idea of optimizing association rules combining with user's score, the steps of optimizing association rules combining with user's score are further discussed in detail below.

3.1 Finding the Valuable Rule

Let D be the transaction database. Transactions in the database include not only the user's transaction information, but also the five-star score information of the purchased product. we represents all sets of goods in a database by

$$I = \{item_1, item_2, item_3, ..., item_l\},$$

where *item* represents a commodity, l represents the total number of kinds of commodities, and $D = \{t_1, t_2, t_3, ..., t_q\}$ is the set of all transaction transactions in the database, which contains items in one or more I, where t represents one piece of transaction data, and q represents the cumulative number of all transactions. Using the association rule algorithm, we can obtain a rule library that contains all the set of rules, i.e.,

$$R = \{r_1, r_2, ..., r_i, ..., r_{|R|}\} = \\ \{X_1 \rightarrow Y_1, X_2 \rightarrow Y_2, ..., X_i \rightarrow Y_i, ..., X_{|R|} \rightarrow Y_{|R|}\}$$

$r = \{X \rightarrow Y\}$ represents a recommended rule, X represents the first half of the rule, and Y represents the recommended result of the rule, $|R|$ is the number of all the rules in the database.

Import: Rule library R and the transactions of the target user X_c

```
Rc=∅                    // Set the initial Rc to an empty set
For each ri in R
  If Xi ⊂ Xc            // The Xi is the precursor of the rule ri
  then Rc = Rc ∪ ri
```

Output: Recommended rules that match the user transactions R_c

Fig. 1. Discovery algorithm for valuable rules.

TC (Target Customer), refers to the customer who needs to recommend it, and can be expressed as

$$TC = (X_c, S_c),$$

where $1 \leq c \leq l, X_c$ represents the product portfolio purchased by the user, S_c represents the user's scoring vector for the corresponding product.

In order to recommend to the target user, you first need to find out all the rules that match the target user, namely the valuable rules.

The pseudo-code of the algorithm is Fig. 1.

3.2 Calculate the Score Value of the Rule

After getting the value rule set R_c, you can find that some rules are the same before, but the recommended results are different. At this point, if randomly selected rules to the user, other rules will be abandoned. All rules in R_c reach the threshold of support and confidence, and all match the target user, so all rules in R_c have the ability to recommend to users. If only one of them is selected and the other rules are discarded, then the important rule information may be lost. How do you make full use of all the information in the rule R_c? At this point, these rules need to be clustered. The goal of personalized recommendation is to recommend the results as satisfactory as possible, so we can gather together the rules with the same recommendation results together to form a new combination of rules. Suppose $Group\_R_k$ is one of the combinations,

$$Group\_R_k = \{X_{k,1} \rightarrow Y_{k,1}, X_{k,2} \rightarrow Y_{k,2}, ..., X_{k,j} \rightarrow Y_{k,j}, ..., X_{k,t} \rightarrow Y_{k,t}\},$$

where $Y_{k,1} = Y_{k,2} = ... = Y_{k,j} = ... = Y_{k,t}$.

That is, all the posterior rules in this rule combination are the same. Using this method, clustering all the recommended rules in the rule result set yields a unique rule clustering result. After dividing the rules into clustering, a group is distinguished by the back elements of the rules. There are several rules within each group, and these rules each support their own recommendation results. At this time, as long as the total score of the recommended results in each recommendation combination is compared, the merits of the recommended results can be analysed. Since all the rules in Rc meet the support

Import: Valuable recommendation rule set R_c

$H_c = \emptyset$ // Set the initial H_c to an empty set

For each r_j *in* R_c // The rating for all items within r_j was initialized as null

For each $item_{j,k}$ *in* r_j

$$S_{j,k} = \frac{\sum_{i=1}^{m} sc_i\_item_{j,k}}{m}$$ // The score value of the k th item in rule r_j was calculated

End for

Output: Rule set H_c with score values

Fig. 2. Score value was calculated for the rule.

threshold, there will be many other users' purchase records supporting the rules, and we refer to these users as the support users of each rule, and then take the mean of the support users as the score value of the within-rule items.

Assuming H_c is a set of recommended rules with scores, it can be expressed as:

$$H_c = \{(X_1, SX_1) \rightarrow (Y_1, SY_1), (X_2, SX_2) \rightarrow (Y_2, SY_2), ..., (X_j, SX_j) \rightarrow (Y_j, SY_j), ..., (X_n, SX_n) \rightarrow (Y_n, SY_n)\}$$

where $(X_j, SX_j) \rightarrow (Y_j, SY_j)$ is Rule j in H_c, X_j is a rule precursor, and can be represented as $X_j = \{item_{j,1}, item_{j,2}, ..., item_{j,k}\}$, SX_j is the score vector of the previous item corresponding item, and can be expressed as

$$SX_j = \{S_{j,1}, S_{j,2}, ..., S_{j,k}\}$$

Y_j is the recommended result of Rule j, and can be expressed as

$$Y_j = \{item_{j,k+1}, item_{j,k+2}, ..., item_{j,n}\},$$

SY_j is the scoring vector of rule recommendation results and can be expressed as $SY_j = \{S_{j,k+1}, S_{j,k+2}, ..., S_{j,n}\}$.

Thus we derive a representation of the rule set Hc with score values served for the target user TC. The specific pseudo-code for calculating the rule score within H_c is shown in Fig. 2.

Suppose that the collection of supporting users for Rule r_j is $SC = \{sc_1, sc_2, \ldots, sc_i, \ldots, sc_m\}$, $sc_i\_item_{j,k}$ represents the score value of user sc_i for the kth item in the rule r_j.

3.3 Target User and Rule Similarity Calculation

Using similar users, you can predict the score of the target users on unpurchased products. Since each rule item in the rule set will partially overlap with the target user, it is feasible to calculate the score similarity between the two, and there is reason to believe that the more similar the rule score is to the target user, the closer the recommended result is to

the actual score of the target user. It is not hard to imagine that the more similar the items in the rules are to those in the target user, the more reliable the recommended results are.

For example, there are two rules $r_1 = \{$(three smile toothbrush, family of three toothbrush cup) $\rightarrow$ Chinese toothpaste$\}$, $r_2 = \{$three smile toothbrush $\rightarrow$ black toothpaste$\}$, at this time, found a customer C has bought a product {three smile toothbrush, family of three toothbrush cup, comfortable good soap}, so in the above two rules support and similarity is the same, we will choose r_1 to recommend, which is to recommend the Chinese toothpaste, because r_1 rules and target users C purchase history project similar degree is higher. Therefore, when judging the similarity of rules and target users, both score similarity and item similarity need to be considered.

The prediction value of each member in the rule combination to the target user is calculated by the similarity size of the rule and the target user, and the final group with the highest prediction score is the optimal recommendation. Here, we do not take into account the factors of the support degree and the confidence degree, but the support degree reflects the occurrence probability of the rule, and the confidence degree reflects the reliability degree of the rule, so it should also be considered in the whole algorithm.

Jkard's coefficient is generally a measure of variability between sets [14]. It is calculated as shown in Eq. 1:

$$sim(A, B) = \frac{|A \cup B|}{|A \cup B|} \tag{1}$$

where A and B are two sets, $|A \cap B|$ represents the number of intersection elements of sets A and B,$|A \cup B|$ represents the number of union elements of all sets A and B. The coefficient is 1 when A is exactly the same as B, and when A and B are completely independent, the coefficient is 0. This paper uses the Jkard coefficient to measure the item similarity measure between the rule precursor and the target user.

Cosine similarity is often used in areas such as information retrieval and statistics, and it is also often used to calculate the size of the user score similarity in collaborative filtering algorithms. However, the similarity in this article is mainly to measure the gap between the rule before the score and the user score, here is more to consider the absolute numerical difference between the two, and the rule results are likely to rule before only one item, then the cosine similarity cannot effectively show the difference between the two. In view of this, Euclidean distance is used as the similarity to the target user.

Assuming that $\overrightarrow{x}$ and $\overrightarrow{y}$ are two vectors in a space, the Euclidean distance between the two vectors can be expressed as $d(x, y) = \sqrt{\sum (x_i - y_i)^2}$,

and the similarity between the two tors can be expressed as formula 2:

$$sim(x, y) = \frac{1}{1 + d(x, y)} \tag{2}$$

Suppose $Group\_H_k$ is a combination of rules in Hc, which can be expressed as:

$$\begin{aligned} Group\_H_k = \{&(X_{k,1}, SX_{k,1}) \rightarrow (Y_{k,1}, SY_{k,1}), (X_{k,2}, SX_{k,2}) \\ &\rightarrow (Y_{k,2}, SY_{k,2}), \ldots, (X_{k,j}, SX_{k,j}) \rightarrow (Y_{k,j}, SY_{k,j}), \ldots, (X_{k,t}, SX_{k,t}) \\ &\rightarrow (Y_{k,t}, SY_{k,t})\} \end{aligned}$$

where $Y_{k,1} = Y_{k,2} = \ldots = Y_{k,j} = \ldots = Y_{k,t}$.

The rule combination consists of the rule t with the same recommendation results, then the score similarity between the target user $TC = (X_c, S_c)$ and the rule j in $Group\_H_k$ can be expressed as:

$$Score\_sim_{k,j} = sim(S_c, SX_{k,j}) = \frac{1}{1 + d(S_c, SX_{k,j})} \tag{3}$$

where S_c represents the target user score vector,$SX_{k,j}$ represents the precursor score vectors of the rule c. And the item similarity of Rule j in the target user TC and the rule combination $Group\_H_k$ can be expressed as:

$$Item\_sim_{k,j} = sim(TC, H_{k,j}) = \frac{|X_c \cap X_k|}{|X_c \cup X_k|} \tag{4}$$

where X_c represents the collection of product categories purchased by the target user TC,X_k is the precursor of the rule $H_{k,j}$. The overall similarity of the two similarity indicators is:

$$sim'(TC, H_{k,j}) = \alpha \times Score\_sim_{k,j} + \beta \times Item\_sim_{k,j} \tag{5}$$

where $\alpha + \beta = 1$, $\alpha \in [0, 1]$, $\beta \in [0, 1][0, 1]$,α and β respectively reflect the score similarity and item similarity on recommended results, can be according to the actual two specific impact on the recommended results to set the size, for example, in the supermarket shopping recommendation, lack of customer for the purchased goods score, customer for goods score largely depends on commodity sales to estimate, so you can use project similarity to calculate the similarity between goods, then you can set a lower value α and a higher value β.

The overall similarity between rules and target users should refer to the value of support and confidence. This paper uses the average value of support and confidence to reflect its influence on the results, so the final similarity obtained is as follows:

$$sim(TC, H_{k,j}) = sim'(TC, H_{k,j}) \times \frac{(conf_j + \sup_j)}{2} \tag{6}$$

where and represents the confidence and support size of the rule, respectively.

3.4 Computing the Optimal Recommendation Results

After obtaining the similarity between each rule and the target user, the total score of the recommended results in each rule combination can be calculated, and the calculation formula is shown in formula 7:

$$P_{TC,Y_k} = \frac{\sum_{j=1}^{t} sim(TC, H_{k,j}) \times SY_{k,j}}{\sum_{j=1}^{t} sim(TC, H_{k,j})} \tag{7}$$

where P_{TC,Y_k} represents the final predictive score of the recommended result of the target user TC for the rule combination $Group\_H_k$, $SY_{k,j}$ represents the score value of the recommended result Y_k within each rule combination,$sim(TC, H_{k,j})$ represents the

Table 1. Sample transaction data.

TID	Record of purchase
T100	I_1, I_3, I_4, I_8, I_9 (2,4,3,1,5)
T200	I_1, I_3, I_9 (4,4,5)
T300	I_2, I_6, I_7, I_9 (3,1,3,5)
T400	I_1, I_6, I_7 (4,2,4)
T500	I_1, I_2, I_4, I_7, I_9 (3,5,5,4,3,4)
T600	I_1, I_7, I_8, I_9 (4,3,4,4)
T700	I_4, I_7, I_8, I_9 (4,5,3,1)
T800	I_1, I_4, I_5, I_7, I_8, I_9 (1,2,1,4,4,3)
T900	I_1, I_2, I_5, I_7 (4,3,2,4)

overall similarity between the rule $H_{k,j}$ and the target user TC, and t represents the total number of all rules within the $Group\_H_k$ rule combination.

In general, the recommended system for the number of target users is not the same, specific to recommend to target users how many goods according to the actual situation, but in the final result sorting, only need to consider the recommended result number between the same rule combination, the highest total score of the one rules group recommended result is the optimal recommendation. It should be pointed out here that the predicted score value obtained by this method cannot represent the predicted score value of the target user on the recommended results, which is only used to compare the merits of the recommended results.

4 A Calculation Example

To better illustrate the working steps of the association rule optimization method combined with user scoring, we explain the various steps of the algorithm with a simple example.

Table 1 is the transaction data of a partial online store item, the data contains a collection of items such as. There are 9 transaction records, and the value after each purchase record is the customer's score on the corresponding product item, with the five grades of 1 to 5 points. Suppose that one user TC has already purchased the product, and the right products were rated as (4,3). So you can recommend a suitable product to it accordingly. Questionnaire design of the variable scale in this study adopts a five-point Likert scale, and the scores are divided into five grades from 1 to 5 points.

Set the minimum support threshold of association rule mining is minSup = 0.3, and the minimum confidence threshold is minConf = 0.6, then the Apriori classical association rule algorithm can finally match the target user TC rules, that is, the rules with recommended value, as shown in Table 2.

According to the classical association rule algorithm, from Table 2, it is not difficult to see that the above 8 rules can be recommended to target users, and the three products

Table 2. A Collection of Value Rules

Rule identification	Valuable rules	Support	Confidence
r_1	I4 → I8	0.33	0.75
r_2	I4, I9 → I8	0.33	0.75
r_3	I4 → I1	0.33	0.75
r_4	I9 → I1	0.56	0.71
r_5	I4, I9 → I1	0.33	0.75
r_6	I4 → I7	0.33	0.75
r_7	I9 → I7	0.56	0.71
r_8	I4, I9 → I7	0.33	0.75

Table 3. Combination of rules with score values.

Set of rules	Valuable rules	Support	Confidence
1	$(I_4,3.00) \rightarrow (I_8,2.67)$	0.33	0.75
	$(I_4,3.00), I_9,3.00) \rightarrow (I_8,2.67)$	0.33	0.75
2	$(I_4,3.00) \rightarrow (I_1,2.00)$	0.33	0.75
	$(I_9,4.20) \rightarrow (I_1,2.80)$	0.56	0.71
	$(I_4,3.00), I_9,4.00) \rightarrow (I_1,2.00)$	0.33	0.75
3	$(I_4,3.33) \rightarrow (I_7,4.00)$	0.33	0.75
	$(I_9,3.6) \rightarrow (I_7,3.40)$	0.56	0.71
	$(I_4,3.33), (I_9,2.67) \rightarrow (I_7,4.00)$	0.33	0.75

I_1, I_7 and I_8 can be pushed to users as the final recommendation results. However, only one product is required to be pushed, the choice is based on the size of the support and confidence. If the support and confidence of rule r_1, r_2, r_3, r_5, r_6, and r_8, is completely equal, and the support of rule r_4 and r_7 is larger than other rules, but less confidence than other rules, so cannot choose a suitable product recommendation to the target user according to support and confidence. And random selection may miss important rule information.

The rules in Table 2 have the ability to recommend target customers, that is, valuable rules. Next, these rules should be classified according to the recommended rules results, and the score value of each rule is calculated. Finally, the rule combination H_c with scores is shown in Table 3.

As can be found from Table 3, the eight rules according to the recommended results can be divided into three different rules combination, each combination of several rules to support their own recommended results, using group decision thinking can make full use of each combination of rules members, that is to say, can use all valuable rules to the target users. As shown in Table 4.

Table 4. Calculates the similarity of the rule results and the target users.

Set of rules	Valuable rules	Support	confidence	Item similarity	Score similarity
1	$I_4 \rightarrow (I_8,2.67)$	0.33	0.75	0.5	0.20
	$I_4, I_9 \rightarrow (I_8,2.67)$	0.33	0.75	1.0	0.50
2	$I_4 \rightarrow (I_1,2.00)$	0.33	0.75	0.5	0.20
	$I_9 \rightarrow (I_1,2.80)$	0.56	0.71	0.5	0.19
	$I_4, I_9 \rightarrow (I_1,2.00)$	0.33	0.75	1.0	0.41
3	$I_4 \rightarrow (I_7,4.00)$	0.33	0.75	0.5	0.55
	$I_9 \rightarrow (I_7,3.40)$	0.56	0.71	0.5	0.20
	$I_4, I_9 \rightarrow (I_7,4.00)$	0.33	0.75	1.0	0.57

Table 5. The composite scores of the recommended results.

Set of rules	Valuable rules	Recommended results score
1	$I_4 \rightarrow (I_8,2.67)$	1.59
	$I_4, I_9 \rightarrow (I_8,2.67)$	
2	$I_4 \rightarrow (I_1,2.00)$	1.75
	$I_9 \rightarrow (I_1,2.80)$	
	$I_4, I_9 \rightarrow (I_1,2.00)$	
3	$I_4 \rightarrow (I_7,4.00)$	3.59
	$I_9 \rightarrow (I_7,3.40)$	
	$I_4, I_9 \rightarrow (I_7,4.00)$	

In order to facilitate the calculation, we take the scoring, so that the weight of the rules in each rule group can be calculated according to formula 6, and then the final scoring results calculated according to formula 7 are shown in Table 5.

If you need to recommend more than two products to users, you can only compare the set of rules with the same number of recommended products. Since this example recommends only one product to users in advance, and the rule group where the recommended result I7 in Table 5 has the highest score, so I7 is the most likely satisfactory product for the target user.

5 Experimental Analysis

The advantages and bad evaluation indexes of general personalized recommendation algorithms basically include several categories of user satisfaction, coverage, prediction accuracy, diversity and novelty. In view of the improvement, this paper is mainly to prove that the combined user score correlation rule optimization method compared with the traditional support and confidence size selection algorithm is more effective, so

mainly consider the recommended user satisfaction and accuracy, and the traditional correlation rules algorithm accuracy discrimination is mainly based on the confidence, and satisfaction is not specific to measure.

The correlation rule optimization method combined with user scoring utilizes the knowledge in the field of user scoring, so the user satisfaction can be specifically defined according to the user score, and the accuracy can be accurately recommended to the users according to whether it is adopted after recommending to the user. The specific expressions of the two judged indicators can be expressed as follows:

Accuracy = the number of usersadopting

recommendation ÷ total number of users recommended by system.

Satisfaction = the number of users adopting

recommendation with high score value ÷ the number of users adopting recommendation.

In order to test the effectiveness of the algorithm, the experimental test data are collected from the MovieLens data set, including the needs of the movie, and the higher the score, the more like the movie, the less interested in the movie, the user selected in the review reviewed at least 20 movies. Experimental according to 80%: 20% ratio randomly generated 6 groups of training set and test set, and then use the traditional association rules Apriori algorithm to get the best recommendation results, using combining the correlation rules optimization method to score the final recommendation results, we will score more than or equal to 4 points users called to achieve satisfactory users.

This paper uses the Java integration tool Eclipse Standard 4.3.2, and the experiments are run in the Win 10 operating system, Core i-4200 U CPU, 2.3 GHz, 4G memory environment. The accuracy and satisfaction comparison of the proposed algorithm and the traditional screening recommendation results, based on support and confidence, are shown in Figs. 3 and 4.

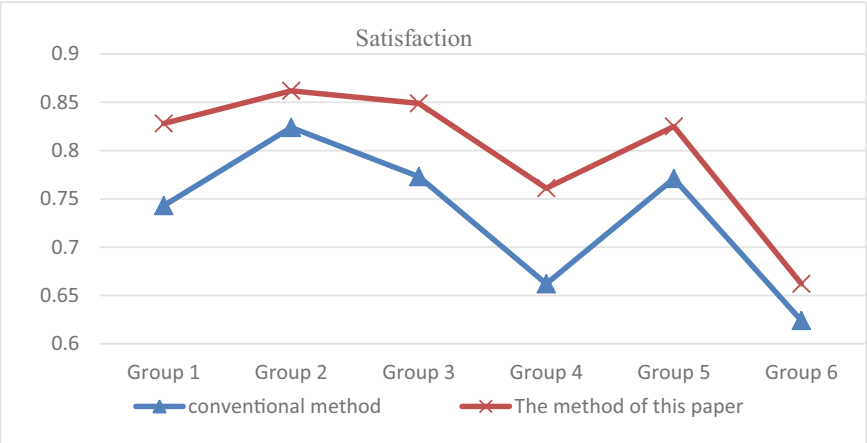

Fig. 3. Comparison of accuracy between the two methods.

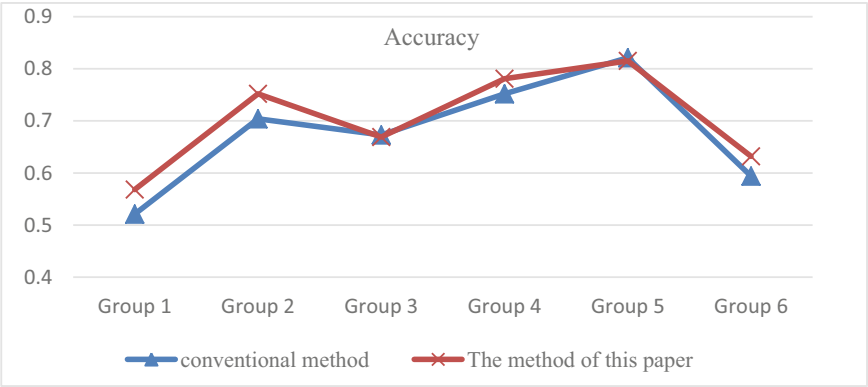

Fig. 4. Comparison of satisfaction between the two methods.

From the comparison of the results of the two methods, we can find that the rule selection method proposed in this paper slightly improves the accuracy of the traditional method, but the user satisfaction is significantly improved, and the recommendation results obtained by the traditional algorithm are relatively low. That is to say, the association rule optimization method combined with user score solves the problem of inconsistency between the same rules before the back, and the support and confidence indicators are difficult to effectively use all rules. At the same time, the improved rule extraction method significantly improves user satisfaction.

6 Conclusion

Since the traditional association rule recommendation generally produces a large number of rule results, and it is difficult to screen the rule results according to the size of support and confidence, there is likely to be conflicting rules, and all the rule knowledge cannot be fully and effectively utilized. This paper proposes a correlation rule optimization method combining user scoring, which can effectively solve the above problems in the database. The method combines the idea of collaborative filtering algorithm to obtain the satisfaction degree that the algorithm can use all the knowledge of the rules in the rule library to obtain the optimal recommendation results more effectively.

This paper explains the working steps of the algorithm in detail by calculating example analysis, and finally uses experimental simulation to verify that the correlation rule optimization method proposed in this paper is more effective than the traditional optimal rules with support and confidence, and the recommended results are more recognized by users.

References

1. Bi J, Zhang Q. Summary of association rule mining algorithms [J]. Chinese Engineering Science, 2005,7 (4): 88–94. In Chinese.

2. Bei-Jun R, Zhu Y Y. Association Rule Clustering Based on Taxonomy Information[J]. Journal of Computer Research & Development, 2004, 41(2):352–359.
3. Ding M, Xiong C. Analysis of oriented association rules in group discussion environment [J]. Computer and Digital Engineering, 2011,39 (6): 36–40. In Chinese.
4. Hu Q, Zhu D, Wu H, Wu L. Review of Intelligent Recommendation Systems Research [J]. Computer System Applications, 2022,31 (04): 47–58. In Chinese.
5. Jiang Y, Shang J, Liu Y. Maximizing customer satisfaction through an online recommendation system: A novel associative classification model[J]. Decision Support Systems, 2010, 48(3): 470–479.
6. Lent B, Swami A N, Widom J. Clustering association rules[C]//IEEE International Conference on Data Engineering. IEEE, 1997:220–231.
7. Liu Y Z, Jiang Y C, Liu X, et al. CSMC: A combination strategy for multi-class classification based on multiple association rules [J]. Knowledge-Based Systems, 2008, 21(8):786–793.
8. Li Q, Qu X, Guan Jun. Similarity measures of association rules with clustering studies [J]. Computer Engineering and Design, 2012,02:745–749. In Chinese.
9. Park D H, Kim H K, Choi I Y, et al. A literature review and classification of recommender systems research[J]. Expert Systems with Applications, 2012, 39 (11):10059–10072.
10. Suzuki E, Ishihara H. Visualizing Discovered Rule Sets with Visual Graphs Based on Compressed Entropy Density[C]//New Directions in Rough Sets, Data Mining, and Granular-Soft Computing, International Workshop, Rsfdgrc'99, Yamaguchi, Japan, November 9–11, 1999, Proceedings. 1999: 414–422.
11. Toivonen B H, Klemettinen M, Ronkainen P, et al. Pruning and Grouping Discovered Association Rules[J]. 2010.
12. Xiang L. Recommended Systems Practice[M]. Beijing: People's Post and Telecommunications Press, 2021. In Chinese.
13. Yu M, He W, Zhou Xuchuan, Cui Mengqi, Wu Keqi, Zhou Wenjie. Recommended systematic review [J]. Computer application: 2022,42 (6): 1–16. In Chinese.
14. Zhang Xiaolin, Fu Yingzi, Chu Peixiao. Application of Jerkard similarity coefficient in recommendation systems[J]. Computer Technology and Development, 2015 (4): 158–161. In Chinese.