

The Improvement of Personalized Recommendation System Based on Sparse Data of Financial Products

Xu Yong, Pi Ying-xin, Xu Jian-bin

Abstract—In the era of big data, innovative marketing of internet finance confronts with new challenges. How to provide users with precise customized financial service, has become a problem of further expansion of internet financial market. Based on this background, the research integrates the ideology of concept hierarchy into the traditional personalized recommendation algorithm, and proposes an improved trick, which is punishing user similarity computation of popular items. And the research integrates the ideology of community mining into recommendation algorithm, making EO algorithm as the basis of community division algorithm, which combines improved index of user similarity and the Q of community discovery to build a new model of community discovery and divide the structure of social network. Optimizing the procedure of recommendation system solves the problems of data sparsity, cold start and system scalability. In the end, selecting the data from the Mint, which is a financial institution in America, demonstrates the improved effect that the two innovative schemes proposed in this research contribute to the accuracy of the recommendation system and the precision of community discovery.

Index Terms—collaborative filtering, community mining, internet finance, personalized recommendation

I. INTRODUCTION

Since 2000, China, starting from scratch, has vigorous development in e-commerce and becomes a leading country gradually. Personalized recommendation technology, as one of applications in the big data and cloud computing, has large development space in China. The research of personalized recommendation began in 2001. Feng Ao et al. developed the recommendation system of information filtering based on Agent, which is effective after information filtering by managing bookmarks on web visitors[1]. In the following years, there are many representative scholars and treatises in the field of personalized recommendation research, including personalized recommendation research in e-commerce and research on recommendation algorithm of collaborative filtering based on project evaluation. In 2007, Peng Yu et al. improved and verified the collaborative filtering algorithm based on the attribute similarity. In 2009, Peng Dekui improved the collaborative filtering algorithm combined with user characteristics and time series, which makes recommendation more effective [3]. Because of the relatively late start of the research about application of personalized recommendation technology, the application of personalized

recommendation in China mostly stays in classified browsing, search engine, and related commodity recommendation. Looking at the current research status in China, there are main problems, such as single recommendation technology, low recommendation accuracy and low continuity[4].

As a new financial model, internet finance began to develop late in China, and its development benefits from the rapid rise of e-commerce. In 2012, Xie Ping and Zou Chuanwei defined the internet financial model for the first time, and proposed that the Internet financial model is a representative modern information technology model [5], which is rising with the development of the Internet. In 2014, Song Mei proposed that Internet finance provided an opportunity and challenge for the development of traditional commercial financial institutions. Internet finance was developed through the optimization and upgrading of services, while traditional banks focused on improving the bank's management of risk by improving the credit rating mechanism [6]. The demand for customized services of financial institutions has promoted the application of personalized recommendation technology in financial institutions. For instance, the Agent model was introduced into commercial financial institutions to provide users with personalized financial products recommendation. Foreign financial institutions have done a lot of research, which shows that customer purchase of financial products and services always presents a highly similar purchase sequence. The purchase sequence of this product and service is called the customer's product acquisition model. According to the theory, foreign scholars Paas and Kuijlen analyze the personalized recommendation service in the financial service field through the Mekken scale and the acquisition mode of financial products, and understand the customers' preference characteristics at different stages in order to predict customer's future behavior bias and purchase motivation [7]. Compared with the focus of foreign scholars, such as cross selling, personalized recommendation, and the acquisition mode of financial products and services, the domestic research on Internet finance mainly stays in the macro level of risk, such as trust, business model and so on. There is less research on the combination of recommendation and Internet finance.

II. THEORETICAL BACKGROUND

Personalized recommendation is one of the important components of data mining technology. It can efficiently filter and analyze data and extract useful information to users. The technology has been applied in movies, books, products, webpages, music, news recommendation, and new media advertising. Among them, the content-based

Xu Yong, School of Economics and Trade, South China University of Technology, Guangzhou, China

Pi Ying-xin, School of Economics and Trade, South China University of Technology, Guangzhou, China

Xu Jian-bin, S.F. Express (Group) Co. Ltd, Shenzhen, China

recommendation algorithm is mainly to extract and filter item information, so most of the theoretical basis, techniques and methods of such recommendation systems are derived from information retrieval theory [8]. Since the content-based personalized recommendation only focuses on the content information instead of the relationship between the item and the user, it can alleviate the cold start problem in the personalized recommendation; the content-based personalized recommendation does not care about the user's comments on the product. Therefore, the sparseness problems that are common in personalized recommendations have little impact on the algorithm.

Personalized recommendation based on collaborative filtering is essentially a recommendation algorithm based on user behavior analysis. It is mainly used to find recommendations by finding users with the same interests or items with the same attributes. The basic idea of the user-based collaborative filtering algorithm [9] is to analyze user groups' historical data, search for user groups that have similar interests with the target users as neighbor users of the target users, and then push the recommendation list for targeted radiation. The project-based collaborative filtering algorithm is actually the push of similar items that are combined with the user's purchase history. Just as you have ever seen thriller, the algorithm will recommend you a thriller movie. You bought the book Introduction to Computers, it will recommend Machine Learning to you.

Personalized recommendation algorithms based on association rules are widely used in the traditional retail sector, in which cases of beer and diapers in the analysis of shopping baskets are widely known. Predecessors have made numerous attempts to apply association rules to personalized recommendation systems. Lin et al. [10] proposed mining algorithms that adjust the minimum scale during the mining process, and to a certain extent, a comparable set of items can be obtained. It solve the shortcomings of the Apriori algorithm, and the accuracy of this algorithm is relatively higher. Domestic scholar Zhang Feng [11] combined association rules with sequence rules to construct an effective recommendation system engine which has a good recommendation effect. However, the algorithm must choose proper support and confidence, otherwise the system will be complicated, redundant and inefficient.

Mixed recommendation is often a combination of multiple recommended algorithms, resulting in a certain combination of rules. The idea of portfolio recommendation has attracted the attention of many scholars. Guo Dongmeng, a domestic scholar, has combined and proposed improvements by combining association rules and collaborative filtering to optimize the personalized web shopping mall recommendation system [12]. The combination recommendation is recognized due to the limitations of various recommendation algorithms, and the combination recommendation technology can always improve the recommendation effect to a certain degree [13]. The combination recommendation model is constructed in various ways. The combination of recommendation algorithms mainly includes the following: weighting results of multiple recommendation algorithms, replacing recommendation strategies according to specific situations, presenting multiple recommendation algorithms, and

combining different recommended data sources's feature, hierarchical superposition recommendation technology, use of new features to add another recommendation technology [14].

III. THE THOUGHT OF SYSTEM RECOMMENDATION AND PROCESS DESIGN

Since the financial industry's products cover a wide range of products, including insurance, trusts, leasing, and bonds, and financial products are generally risky, time-sensitive, and low-frequency, the traditional commodity recommendation technology is not sufficiently accurate in the application of Internet finance. Therefore, this study has improved the research on the problems that appear in the traditional recommendation system.

For the problem of data sparseness, this study uses the idea of concept layering to classify financial products to form financial product clusters with different attributes. Given the financial product i , the result of the financial product classification is I , and the user is U . The difference from the traditional user-based collaborative filtering recommendation algorithm is that the user and the financial product are both used as objects to perform similarity calculations. That is, the U/I matrix is formed to calculate similarity, and users and specific products are not directly calculated. The relationship between i . And use the log function to punish the hot items that users are concerned about, and in this way to improve the effect of finding neighboring users. Afterwards, the concept of community mining is introduced, and the user similarity index is integrated with the indicators of community division, so as to build a new type of community division model, and the user is divided into several sub-communities. The recommendation algorithm based on association rules is introduced into the sub-divided society for finance product's recommendation.

A. The improvement of recommendation algorithm based on content

Currently, the personalized recommendation algorithm based on content is one of the methods to solve the cold start problem of the recommendation system, because it just needs to calculate the similarity between the attributes of any two items or contents to recommend the similar content. The information structure of financial products is mainly composed of the amount of investment, the annualized returns, the term of financial management, the mode of interest rate and the risk rating. Therefore, the first step of calculating the similarity between the items is to quantify the attribute of the project, and to standardize the attributes of each item.

The most typical and effective method of data standardization is the 0-1 standardization. It defines x_n as the standardized value of the n -th attribute of financial products [15]:

$$x_n = \frac{x_n - \min}{\max - \min} \quad (1)$$

We generally calculate the similarity between two items by Euclidean distance. Therefore, the calculation of similarity between financial products uses the Euclidean distance. Assuming that a product has n attributes, any two

financial products can be expressed in the form of vector, such as $X\{x_1, x_2, \dots, x_n\}$ and $Y\{y_1, y_2, \dots, y_n\}$, where x_n represents the n -th attribute of product X , y_n represents the n -th attribute of product Y . The distance between the two products is [15]:

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

The Euclidean distance algorithm has great time complexity. Assuming that there are N items and each item is described by M attributes, the complexity of the algorithm is $O(Nm^2)$. Therefore, in the actual commercial application, in order to improve the operational efficiency of the recommendation system, the recommendation algorithm is optimized by establishing the inverted list of key words - items.

B. The improvement of Collaborative Filtering Recommendation algorithm Based on User

Usually, the collaborative filtering recommendation algorithm based on user directly calculates the relationship between users' financial products i , and finally forms the similarity matrix U/i . Due to the amount of financial products is too large, the data matrix between the user and the financial product is sparse. It is difficult to calculate the accurate similarity between the users. According to the idea of dimensionality reduction, the financial products i are classified into 5 categories, which are $\{I_1, I_2, I_3, I_4, I_5\}$ respectively. Finally, they form the similarity matrix U/I , which is not sparse.

The collaborative filtering algorithm is based on user's behavior to calculate the similarity between users. In the process of calculating users' similarity, the data sparsity can be effectively reduced by concept stratification. Given a user U_i , $N(U_i)$ represents a set of items with positive feedback from user U_i . Through cosine similarity calculation method, we can calculate the interest similarity of U_1 and U_2 [15]:

$$w_{U_1 U_2} = \frac{|N(U_1) \cap N(U_2)|}{\sqrt{|N(U_1)| |N(U_2)|}} \quad (3)$$

The time complexity of this method is $O(|U|*|U|)$, so it will reduce the efficiency of the system when the user data in the system is too large. Due to the low amount of users concerned, it is often difficult for most users to determine the relationship between the two products. In order to solve this problem, the previous research is mainly to establish the inverted list of the goods to the user. The idea of this study is to classify the items and study the relationship between the user and the product class, which can not only solve the problems mentioned above, but also solve the problem of sparse.

After obtaining the interest similarity between the users, the UserCF algorithm recommends the project I to the user, that the other K users who are most similar to his interest are concerned with. The following formula can be used to calculate the degree of interest of user U_1 on the project class I : [15]:

$$p(U_1, I) = \sum_{U_2 \in S(U_1, K) \cap N(I)} w_{U_1 U_2} r_{U_2 I} \quad (4)$$

$S(U_1, K)$ represents the K users who have closest interest with the user's U_1 interest. $N(I)$ refers to the set of users who

have behavior with project I . $w_{U_1 U_2}$ is the interest similarity between the user U_1 and the user U_2 . $r_{U_2 I}$ is the interest of the user U_2 on the item I . For this study applies implicit back feed for the data of the user's behavior, make all $r_{U_2 I} = 1$.

The above formula doesn't consider the impact of popular projects on the calculation process, such as the impact of popular investment projects in this study on interest similarity. The great majority chinese have seen the TV drama of the journey to the west. So if two users have seen the journey to the west, the conclusion that they all prefer mythical TV dramas can't be drawn. But two users have bought "machine learning", and they can draw a big similarity in interest because these books tend to be purchased only by researchers in the field. And non-professionals tend not to have too much attention on them. The group concerning unpopular goods is often small. The smaller the group is, the more it can discover its interest preference and achieve precise marketing. Therefore, in order to punish the impact of hot commodities, this study uses the log function to punish the popular items in the lists of users' interest. Using this way reduces the impact of the popular items on the user's similarity in the calculation process. Therefore, this study will use the following formula to calculate user similarity.

$$w_{uv} = \frac{\sum_{i \in N(u) \cap N(v)} \frac{1}{\log |N(i)|}}{\sqrt{|N(u)| |N(v)|}} \quad (5)$$

C. The improvement of recommendation algorithm based on community mining

Newman algorithm is basic algorithm of most community mining algorithms, and it is representative. Aiming at the design of financial products and recommender systems and combining with Newman, EO algorithm and collaborative filtering algorithm based on user, a new community mining algorithm is proposed in this study. Newman fast algorithm is an agglomerative Algorithms based on greedy strategy. Given the initial network has n communities and a node is a community. Given the initial value of e_{ij} and a_i , as the following [16]:

$$e_{ij} = \begin{cases} 1/2m, & \text{when link between } i \text{ and } j \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

$$a_i = k_i / 2m \quad (7)$$

k_i is the degree of node i , and m is the total number of sides of the community. In the process of community partition calculation, through continuous iteration and merging, the increment of the Q value after the merger is ΔQ . The calculation is based on the maximum combination of ΔQ . The ΔQ is as follows:

$$\Delta Q = 2(e_{ij} - a_i a_j) \quad (8)$$

The merging process of a community is carried out in the direction of the largest or smallest ΔQ . After each merger, e_{ij} , a_i , a_j are updated. This process is iterated until the $N-1$ is merged, and the whole community will be merged into a large community. After the whole algorithm is completed, the tree type community structure is formed, and the connection segments of the tree graph are divided into different sub

communities, and the best community structure can be obtained by selecting the biggest Q value sub community. The EO algorithm is improved on the basis of Newman algorithm, which has a new definition of the Q value and defines the q value of the local module from the global module Q value. The q value is a part of the data representation of the community structure of the Q value, and the q value is generally defined as [17]:

$$\lambda_i = e_{g,i} - a_g \quad (9)$$

$e_{g,i}$ is the proportion of sides of the vertex i and the community g to the all sides. a_g refers to the proportion of all sides of the community g to the entire network edge, and the q value of the local module represents the degree of membership of node i to the community g. EO algorithm can automatically determine the number of communities according to the optimization of q value without realizing the number of communities.

The improved user similarity is used to partition the community. Given value is β and β is used to measure the relationship between nodes.

$$\beta = \frac{1}{2}w + \frac{1}{2}\lambda \quad (10)$$

Given a threshold range, only the nodes that meet the setting threshold can be classified into a community. This study also takes the above methods to realize the division of the subdivision of the financial users, which is called S-EO algorithm.

IV. DATA PROCESSING AND DESIGN

A. Data processing

The data selected this study comes from the data of the financial products purchased by the users of the well-known foreign financial platform Mint's website from July to October 2015. It contains information on the transaction of 384 kinds of financial products by 10,000 users. After excluding users' incomplete information, data of financial products which did not cover the time from July to October, and data which is complicated and unsuitable for the recommendation system to process, such as some financial product portfolios and other data information, and at the same time choose the purchase data which contains three kinds of financial products, and the frequency of which is more than once a month. The transaction data of 870 users for 172 kinds of financial products are screened out, a total of 3,491 transaction data.

The data structure of financial products includes the following fields: {investment line, annualized rate of return, period of financial management, method of interest calculation, amount of starting investment, risk rating}. The user data structure includes the following fields: {Customer ID, customer age, gender, geographic area, income, marriage, whether or not there is a child, whether or not there is a property, whether there is a car, VIP customer type, financial product purchased, amount purchased}. Combining the research needs of this study, the classification criteria of the website (by entering four screening dimensions to find out whether there are related products) and the actual investment situation, the financial products are divided into the following seven categories: high risk, long-term, super-return, T+1; high risk, long-term, over-return, T+N; high-risk, short-term, over-return, T+1; high-risk, short-term, over-return, T+N;

low-risk, long-term, general return, T+N; low-risk long-term, over-return, T+N; low-risk, short-term, general income, T+1. In an offline experiment that measures personalized recommendation algorithms, the data set is divided into two parts: training set and test set. For each customer, the last financial product of its transaction is selected as the target of prediction, and other products are used to build the similarity model as a training set. In the end, if the user's last purchased product appears in the recommendation list, the recommendation is successful.

B. Evaluation index

In the segmentation result of social network, the accuracy of community classification NMI[18] is an important evaluation index and can be defined as:

$$NMI = \frac{-2 \sum_{i,j} N_{ij} \log \left(\frac{N_{ij} N}{N_i N_j} \right)}{\sum_i N_i \log \left(\frac{N_i}{N} \right) + \sum_j N_j \log \left(\frac{N_j}{N} \right)} \quad (11)$$

N is the matrix obtained after two different algorithms are calculated. The NMI value ranges from [0, 1], and the closer the NMI value is to 1, the more accurate the network community is. N_{ij} represents the number of nodes contained in the communities c_i and c_j at the same time; N_i represents the sum of the elements of the i-th row of N. The NMI value and Q value are used as two evaluation indicators for community mining to comprehensively analyze the accuracy and effect of community classification.

The recommendation system provides a personalized recommendation service for the user, and ultimately returns to the user a product recommendation list that the system believes the user is interested in. This recommendation method is called Top-N recommendation. The evaluation of Top-N recommendation effect is generally measured by precision and recall. In order to comprehensively evaluate the effectiveness of TopN recommendations, a set of precision and recall rates will generally be calculated with different recommended list lengths N, and then a correlation curve will be drawn.

Given $R(u)$, a list of recommendation is generated from the training set, $T(u)$ is a list of user actions on the test set. The recall of recommended results can be expressed as [15]:

$$Precision = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (12)$$

The accuracy of recommendation results can be expressed as [15]:

$$Recall = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \quad (13)$$

Because the improvement of the precision may bring about a decrease in the recall, and both indicators are critical to the evaluation of the recommended algorithm, so in the past, scholars introduced a comprehensive evaluation index F in the study, called the reconciliation rate, generally The higher the reconciliation rate is, the better the recommendation effect is. The harmonic ratio F is defined as follows [15]:

$$F = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \quad (14)$$

With the off-line experiment, we can also obtain a variety of evaluation indicators. The user's interest is varied. Although the long-term behavior is consistent, the

recommendation list cannot only reflect the user's single interest. Diversity and similarity are often corresponding to each other. Given $s(i, j)$ defines the degree of similarity between items i and j , and its range is $[0, 1]$, then the diversity of user u 's recommendation list $R(u)$ is defined as follows:

$$Diversity = 1 - \frac{\sum_{i, j \in R(u), i \neq j} s(i, j)}{\frac{1}{2}|R(u)|(|R(u)|-1)} \quad (15)$$

The diversity of the recommended system can then be defined as the average of the diversity of all users' recommended lists, which is expressed as follows:

$$Diversity = \frac{1}{|U|} \sum_{u \in U} Diversity(R(u)) \quad (16)$$

This study selects the evaluation indicators of the recommendation system based on precision, recall, F value and diversity. These indicators can be optimized through off-line methods. At the same time, the evaluation dimension is often considered when evaluating the index evaluation recommendation system. The evaluation dimension is generally divided into a user dimension, an item dimension, and a time dimension. Different indicators have different performance in different dimensions, and the performance of the recommendation system also has different performance in different dimensions. However, this study considers the limitations of the data structure and does not consider the impact of the evaluation dimension.

V. ANALYSIS AND DISCUSSION OF EXPERIMENTAL RESULTS

A. Classification result and analysis of improvement result of user similarity algorithm

In the previous article, we have mentioned that this research improves the user's similarity calculation by punishing the popular items. The purpose is to improve the recommendation efficiency of the recommendation system and solve the data sparsity. In the traditional recommendation algorithm, the recommendation algorithm based on content is CBF, the collaborative filtering recommendation algorithm based on user is UserCF, the collaborative filtering recommendation algorithm based on project is ItemCF, and the recommendation algorithm based on association rules is AF. The similarity calculation proposed in this study is I-n. The evaluation indexes include accuracy (Precision), recall (Recall), harmonization (F), running time (Time), and TOP-N recommended N value to be set to 20. The analysis results are as follows:

Table 1 Analysis of algorithm improvement effect

	Precision	Recall	F	Time (ms)
CBF	0.564	0.323	0.41076	203
I-CBF	0.577	0.326	0.416616	179
UserCF	0.578	0.304	0.39844	196
I-UserCF	0.591	0.312	0.408399	172
ItemCF	0.487	0.298	0.369748	212
I-ItemCF	0.509	0.303	0.379869	182
AF	0.593	0.309	0.40629	217
I-AF	0.634	0.319	0.424441	186

The above results show that the accuracy, recall and harmonization of the proposed algorithms are improved by using the improved algorithm and the classification standards of this study, and the proposed algorithm can greatly shorten

the computing time and improve the scalability of the system. In a word, the idea of user similarity improvement and classification proposed in this study has improved the effect of the recommendation system, and can solve the problem of data sparsity and system scalability to a certain extent. It also shows that project similarity w is an effective index for user community partition.

B. Analysis of the result of user community detection

In order to facilitate the display of user community detection map, the transaction data of 870 users and 172 kinds of financial products are randomly selected, and 100 users are selected as network nodes to draw the social network structure map. The GN algorithm, Newman algorithm, EO algorithm and S-EO algorithm are run separately, and the results are as follows:

Table 2 The result of community detection

Algorithm Name	GN	Newman	EO	S-EO
Community Number	5	4	6	7
Qvalue	0.636	0.677	0.682	0.767
NMI Value	0.74	0.87	0.84	0.92

The data table shows that the GN algorithm divides the network into 5 communities, the Q value is 0.636, the NMI precision is 0.74; the Newman algorithm divides the network into 4 communities, the Q value is 0.677, the NMI precision is 0.87; the EO algorithm divides the network into 6 communities, Q is 0.682, and NMI precision is 0.84; The improved S-EO algorithm divides the network structure into 7 communities, the Q value is 0.767, and the NMI precision is 0.92. By comparing the above data, we can find that the improved S-EO algorithm shows better computational accuracy, and the Q value is also in a higher ranking position in the 4 algorithms. At the same time, the Newman algorithm also has high computing precision, because the operation effect is determined by the objective factors such as the order of the access node of the algorithm, so the number of community division often exists uncertainty. From the perspective of the number of communities, the improved community partition algorithm can subdivide social networks into more groups, which makes it easier to achieve precise recommendation. In order to understand the community structure of the four algorithms more intuitively, the visualization software Ucinet was applied in this study.

The network structure diagram of GN algorithm consists of 100 nodes, 163 sides and 5 sub communities. As shown in the following figure:

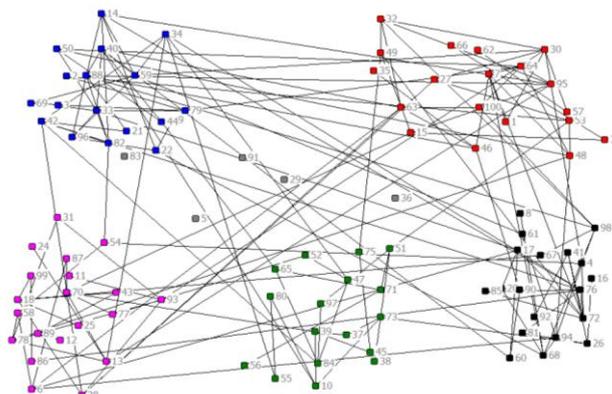


Figure 1 Network structure diagram of GN algorithm

The network structure diagram of Newman algorithm consists of 100 nodes, 131 sides and 4 sub communities. As shown in the following figure:

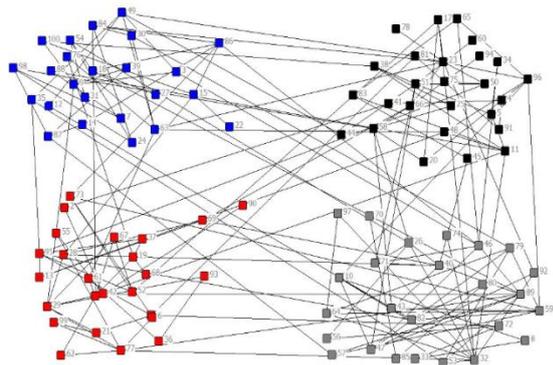


Figure 2 Network structure diagram of Newman algorithm

The network structure diagram of EO algorithm consists of 100 nodes, 178 sides and 6 sub communities. As shown in the following figure:

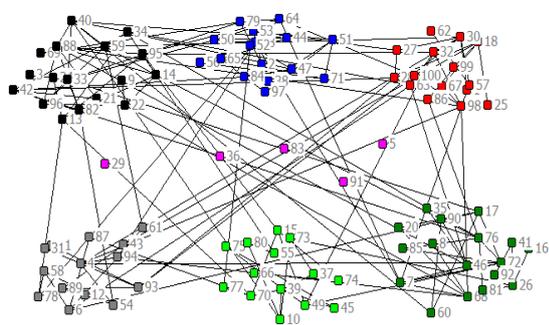


Figure 3 Network structure diagram of EO algorithm

The network structure diagram of S-EO algorithm consists of 100 nodes, 218 sides and 7 sub communities. As shown in the following figure:

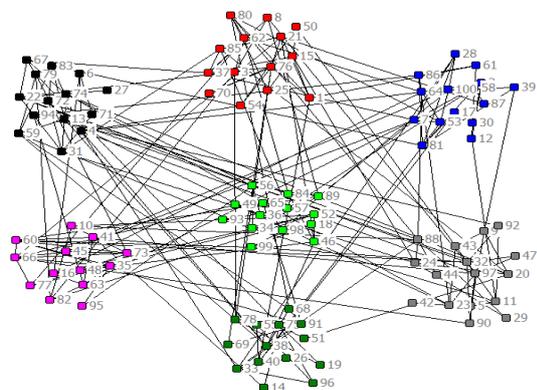


Figure 4 Network structure diagram of S-EO algorithm

In the social network structure diagram, it is found that the four algorithms can clearly subdivide the different subcommunities, but only the S-EO algorithm can effectively divide the 100 network nodes, while the EO algorithm and the GN algorithm have some nodes that can't integrate into the community. In addition, in the community structure divided by S-EO algorithm, nodes and nodes are more closely linked, and the connections between different sub communities are relatively sparse. Therefore, the S-EO algorithm can make more efficient community division of the network, and provide a prerequisite for the next step in the recommendation system to implement the personalized recommendation of the subdivision financial users.

In this study, S-EO algorithm is used to classify 870 users. Finally, a social network containing 870 nodes, 7

communities and 1923 sides is formed, with a Q value of 0.761 and a NMI of 0.93. Then, the personalized recommendation algorithm based on association rules was applied to the recommendation of financial products in 7 communities. The average value of each data is taken as the evaluation index for the recommendation effect of the social network.

C. Analysis of personalized recommendation algorithm for subdivided users

According to the experimental design, the last item of the customer's behavior is divided into the object of recommendation system prediction, so the size of each user's test set is 1. Personalized recommendation based on association rules is used as the main recommendation method. In TOP-N recommendation, the value of N is {5, 10, 15, 20, 40, 80}. The Q value of the community partition is selected as 0.761, and the F value harmonic ratio and diversity are selected as the evaluation index of the recommendation system. Recommendation algorithm after introducing the idea of association mining and 4 kinds of improved algorithms: I-CBF, I-UserCF, I-ItemCF and I-AF were compared and analyzed.

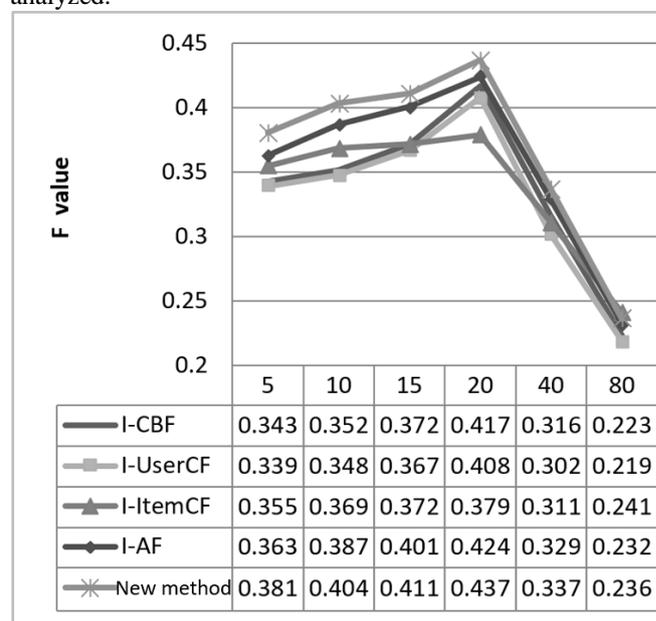


Figure 5 Comparison diagram of F value (algorithm harmonization)

Compared with the four recommendation algorithms, the method adopted in this study (New method) has improvement of a certain degree in reconciliation. This shows that the introduction of community mining can identify the user interest community better and enhance the effectiveness of the recommendation system. In addition, when the N value is 20, the overall recommendation is the best.

Another evaluation index can be obtained through offline experiments: diversity. If the recommendation list is only a single user's interest, then the recommendation list fails. Diversity and similarity are often corresponding. The analysis of diversity results is as follows:

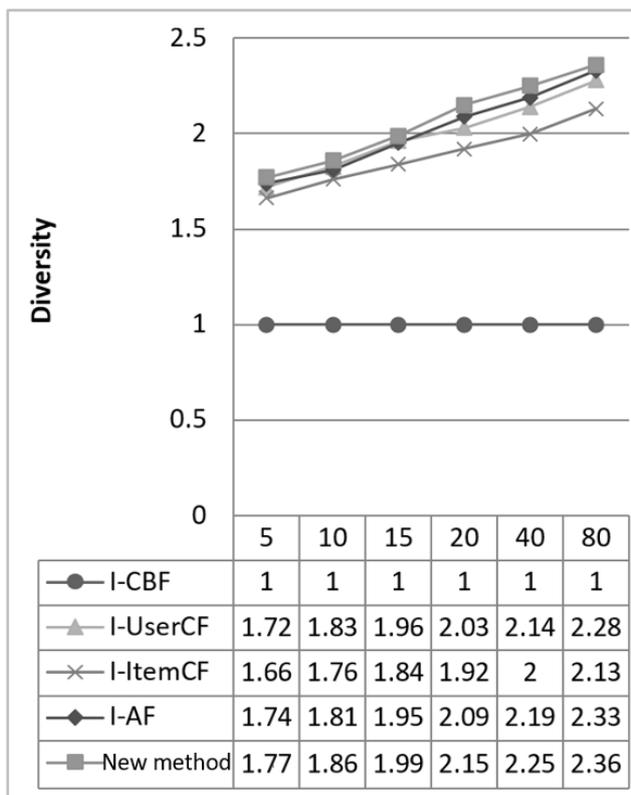


Figure 6 Comparison of diversity among five algorithms

As can be seen from the above diagram, the diversity of various algorithms shows an increasing trend on the whole. Because the greater the value of N values, the more abundant the recommended products will be. The recommended methods proposed in this study are higher in diversity than in other methods. The order from high to low is the method adopted in this study(New method), I-AF, I-UserCF, I-ItemCF, and I-CBF. Personalized recommendation method based on content is calculated in the same category, so the diversity is 1.

VI. SUMMARY AND OUTLOOK

This study elaborated the necessity of developing personalized financial services in the financial field, and based on previous research, summarized the research status of the personalized recommendation system and internet finance at home and abroad. Based on previous research, the key issues to be solved for financial products are proposed, including data sparsity, cold start, and the optimal division of communities, etc. The evaluation indexes of the recommendation method are summarized, and several evaluation indicators suitable for off-line experiments are taken as the evaluation index in this study. Finally, based on the data of the foreign Mint financial product investment platform, we design the experimental process. Using the user similarity improvement algorithm and the classification criteria of this study, the accuracy of the recommendation system can be greatly improved, indicating that the similarity index w is an effective measure. Then by running GN algorithm, Newman algorithm, EO algorithm and S-EO algorithm, the effects of four kinds of community mining algorithms are analyzed. It is found that the S-EO algorithm has higher precision in the classification of network communities, and validates algorithm this study proposed has a good recommendation effect.

Due to the limited experimental conditions and individual research level, this study has the following deficiencies: the index of user similarity w and the value of q are taken as the basis for construction but the weights of both are not considered, and the difference between the two weights may effects the accuracy of community division. The volume of financial data is huge and this study does not propose a solution to process the huge data volume. The customer segmentation method proposed in this study does not consider the offset problem of user interest points, especially for the financial sector. For example, the user's investment preferences will change as time goes by.

REFERENCES

- [1] Feng Ao, Liu Bin, Lu Zengxiang et al, Open Bookmark -- information filtering system based on Agent, Journal of Tsinghua University(Science and Technology), vol. 2,2001,pp.85-88.
- [2] Peng Yu, Cheng Xiaoping, Item-based collaborative filtering algorithm based on attribute similarity, Computer Engineering and Applications, vol. 14. 2007, pp. 144-147.
- [3] Peng Dewei, Hu Bin, A collaborative filtering recommendation algorithm based on user characteristics and time, Journal of Wuhan University of Technology, vol. 3, 2009, pp. 24-28.
- [4] Ouxiaoheng. Research on collaborative filtering based recommendation algorithm. Maritime Affairs University Of Dalian,2015,pp.36-37.
- [5] Xie Ping, Zou Chuanwei, Research on Internet financial model, Journal of Financial Research,vol. 12,2012,pp.11-22.
- [6] Song mei, Analysis of the influence of Internet financial mode on traditional banks, Money China,vol. 3,2014,pp.30.
- [7] Jiawei Han, Jian Pei. Mining Frequent Paterns by Pattern-Growth: Methodology and Implications, ACM SIGKDD Explorations (Special Issue on Scaleble Data Mining Algorithms) ,vol. 2,2000,pp.14-20.
- [8] Song ruiping, Research on hybrid recommendation algorithm, Lanzhou University, 2014,pp.21-22.
- [9] Hofmann T. Latent semantic models for collaborative filtering. Acm Transactions on Information Systems, 2013, 22(1):89-115.
- [10] Lin W, Alvarez SA and Ruiz C. Efficient adaptive-support association rule mining for recommender systems. Data Mining and Knowledge Discovery, 2002, 6(1): 83-105.
- [11] Zhang Feng, a regular friend, Yi Yang, Rule based e-commerce recommendation system model and Implementation, Computer Integrated Manufacturing Systems, vol. 8,2004,pp.898-902.
- [12] Guo dongmeng, Research on Recommendation Algorithm in personalized recommendation system, Beijing Jiaotong University,2017.
- [13] Wozniak Michal, Grana Manuel, Corchado Emilio. A survey of multiple classifier systems as hybrid systems [J] · Information Fusion, vol. 3,2014,pp.3-17.
- [14] Zhang zhiqi. Research on matrix decomposition recommendation algorithm based on user relationship , Beijing Jiaotong University,2016.
- [15] Xiang liang, Practice of Recommendation System , Posts and Telecommunications Press,Beijing,2012.
- [16] Ling xiaowei, Research on multi-relational social network mining based on subspace clustering , University of Electronic Science and Technology of China,2016.
- [17] O. Brauer, H. A. Camargo, M. Socolovsky. Newman-Janis Algorithm Revisited [J]. International Journal of Theoretical Physics, vol. 54,2015,pp.302-314.
- [18] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms [J]. Physial Review E, vol. 78, 2008,pp.6-12.

Xu Yong Associate professor, consultant of China electronic finance industry alliance, focusing on Internet finance research.

Pi Ying-xin Graduate student of South China University of Technology, focuses on Internet finance.

Xu Jianbin Project manager of Shenzhen Shun Feng company, focuses on Internet finance.