2019 2nd International Conference on Safety Produce Informatization (IICSPI)

The Application Research of Customer Segmentation Model in Bank Financial Marketing

Yongxiang Feng
School of Data Science and Application
Inner Mongolia University of Technology
China
yxfengyx@126.com

Xiaoxin Wang
School of Data Science and Application
Inner Mongolia University of Technology
China
Wangxiaoxin5683@163.com

Leixiao Li
School of Data Science and Application
Inner Mongolia University of Technology
China
llxhappy@126.com

***Abstract*—With the rapid development of information technology and internet finance, RFM model technology has been widely used in banking financial services. The traditional financial services industry has gradually changed from product center to customer center. As customer transaction data continues to grow in the database, it is urgent to analyze the transaction information with the high efficiency big data analysis technology. Therefore, an RFM model suitable for financial customers has been established, and this model has been applied to the financial product recommendation guidance system. It can greatly improve the quality of customer marketing services in the banking industry, and can effectively reduce operating costs.**

Keywords—RFM model, customer segmentation, recommendation guidance

I. INTRODUCTION

At the beginning of the 21st century, China joined the WTO, which promoted the further development of the domestic financial market pattern. Financial markets are enjoying a new boom [1].The prosperity of financial market makes between traditional banking and all kinds of new type of financial services industry to customers increasingly fierce [2]. The traditional banking industry relying on the inherent mechanism consists of various branches of the passive customer marketing, which can't satisfy the needs of customers and can't provide customers with satisfactory service [3]. In recent years, wealth management business has been developing rapidly in traditional banking. As a result, many Banks have turned their wealth management businesses into key businesses. The development of personal finance business has promoted the marketing reform of traditional banking industry. As the core to the customer's core market transformation, financial business marketing services are also changes in the traditional Banks.t has become an undisputed consensus to provide customers with appropriate level services and to provide high-quality, efficient and targeted banking financial marketing recommendations to customers [4].

II. RELATED THEORIES

A. Application of RFM Model for Financial Clients

Based on the analysis of the transaction information behavior of financial clients, the RFM model of customer value was created. The model is mainly applied to the classification of customers. According to the system management of customer classification information, the model can reach the purpose of formulating targeted marketing strategy for the marketing personnel. The system is recommended by the customers in accordance with the grade to achieve the purpose of marketing guidance. The RFM model does not require a specific value calculation for the customer lifetime value (CLV), so it is more practical [5].

Three of RFM variables in determining the specific calculation method, which usually standardized processing. After processing, the data will be calculated or clustered [6,7]. The concrete process will be conducted in the next chapter in detail. First of all, the information of original data prepossessing to get score, then the rating information for K - means clustering, finally analyze the clustering results and get the customer classification information, this is the way of FRM model building.

The RFM model on data processing, operate more easily than other models, the results of the analysis is also easier to understand [8]. Based on 144109 transaction records of 3,7170 clients, this paper deals with the transaction data of financial clients of commercial Banks. According to RFM model, data prepossessing is conducted and customer segmentation model is established. The model uses the distribution of R, F and M score values to classify customers, and finally analyzes the customers by using the results obtained by clustering.

B. RFM Model in Financial Customer Segmentation

Although RFM model is a common method for customer segmentation, RFM indicators need to be optimized and improved according to specific conditions. In order to apply to the bank financial clients transaction information data analysis [9]. If established in accordance with the traditional model of RFM indicators will appear many problems, for example, the last trading date only statistical customer with the number of days interval, testing benchmark dates will not be able to analyze the bank financial trading time span of the customer, unable to distinguish between old and new customers at the same time; For example, a customer in a certain period of frequent to the operation of the open to redeem to buy financial products, trading frequency index will continue to accumulate, unable to accurately analyze the customer segmentation frequency standards. Since the amount of investment is also an accumulative value, it will result in a very large accumulative value in the transaction of frequent redemption and repurchase, which is obviously unreasonable. So the traditional RFM model index cannot apply to bank financing of customer segmentation. To this end, the paper puts forward using customer enclosed products trading information as data base, and quarterly average investment according to the customer for data analysis, and then established is suitable for the financial RFM model of customer segmentation. The traditional RFM index and improved RFM index for such as shown in Table I.

TABLE I. THE TABLE OF COMPARISON OF INDEX MODEL

Model indexes	Traditional RFM index	In this article, the financial customer RFM index
R (Proximity)	The time interval between the customer's last trading distance test point	For the first time in a certain period (two years), the time interval between the distance analysis points of the closed financial products is purchased. The time interval between the distance analysis points of the closed financial products is the last time the customer purchases the closed financial products.
F (Frequency)	The total number of times a client trades in a period of time.	The average number of times (quarterly frequency) of each type of financial products purchased by customers within a certain period (two years).
M (Monetary)	The total amount of money that the customer deals for a period of time.	During the period (two years), the customer will purchase the amount of the closed financial products every quarter.

III. CONSTRUCTION AND SYSTEM APPLICATION OF FINANCIAL CUSTOMER SEGMENTATION MODEL

A. Research Content

Firstly, the improved RFM model is used to model the customer segmentation, and then the Oracle relational database is used to store the original data of the model. The system uses the Linux language editing system script to set the data transmission of the Oracle relational database and the HDFS distributed file system. The system uses Scoop data transmission technology as the transmission bridge between two databases. Secondly, the system uses the machine learning library in Spark On Yarn platform to optimize k-means clustering technology. Through the combination of data and data mining technology, the customer data can be analyzed effectively and the valuable information of the customer can be deeply mined. Finally in the system through the clustering results of customer segmentation model of display, and analyze the content classification information. To bring more opportunities recommended guidance for traditional banking, these opportunities are buried in the customer transaction behavior.

B. Preparation of Data

(1) Data Acquisition

The initial data is statistics from the bank's financial trading system, the system after a unified text format specification, system to save the data text, thus ensuring the data consistency. The bank's financial transactions system will process the detailed transaction data of the customer, the necessary system user data and customer information data, and then enter into the system.

(2) Data Preprocessing

1) Data Cleaning

The main purpose of data cleaning is to deal with invalid data in the data, and the scope of invalid data is directly related to the selection of the model and the property of the transaction data itself.

The invalid data caused by the selection of the model in this article mainly includes: the transaction attributes required for non-RFM models can be removed directly. At the same time, the missing data corresponding to the three values of proximity, frequency and value should be filled and processed.

Invalid data in transaction information is mainly includes: the failed transaction in the transaction data, the remote data and products of distance model time, and the data of non-closed type.

2) Data Specification

The primary purpose of the data attribute specification is to delete the attributes that are not related to the model in the original data. For example, the entry teller, customer name and product dividend method in the transaction information. Select the valid attribute for the data set through the data attribute specification: customer number, transaction confirmation date, transaction confirmation amount and transaction number.

3) Data Conversion

The main purpose of data conversion is to unify the different measurement units in raw data. Because transaction information data is processed, the transaction date, transaction frequency and transaction amount are different, standardization is required before the model construction is carried out.

(a) Standardized processing of data in the near degree

In this paper, the near degree values of RFM model are standardized. The process is to record the first transaction date Rf and the last transaction date Rl, i.e. the date of the transaction after the first transaction and the last transaction confirmed by the customer.

The data in the paper comes from the real data provided by a city commercial bank. The data provided from the beginning of 2011 to the end of mid-April 2017. This paper selects the data from April 2015 to March 2017 as the initial data set of the model, and the observation date is selected as the last day of April 2017.

During April 2015 to March 2015 customer transaction data according to the trading time for positive sequence arrangement, which can be counted in the time range of the customer's trading for the first time the date and time at the end of the transaction date. The time interval of this paper is two years, and there are eight quarters in total, which can be divided into eight scores to evaluate the difference, which fully reflects the difference of the importance of proximity in predicting customer transaction behavior. Therefore, according to the customer's transaction confirmation date, the distance from the observation time is 1, and the recent rule of 8 is graded. Furthermore, the corresponding values of Rf and Rl are evaluated as Rfs and Rls. The specific Rf and Rl processes are shown in Table II.

TABLE II. THE TABLE OF RECENCY DATA NORMALIZATION.

Start Date	End Date	Score Evaluation	Customer Share (%)
2015-04-01	2015-06-30	1	1.821
2015-07-01	2015-09-30	2	5.736
2015-10-01	2015-12-31	3	7.474
2016-01-01	2016-03-31	4	13.352
2016-04-01	2016-06-30	5	14.563
2016-07-01	2016-09-30	6	16.083
2016-10-01	2016-12-31	7	26.478
2017-01-01	2017-03-31	8	14.493

(b) Standardized processing of frequency data

Average customer purchase frequency Fq is a customer during April 2015 to March 2017, the number of customers per quarter investment enclosed products, is the customer in this transaction successful quarter average number within two years. The standardized treatment is shown in Eq. 1. Which Fq for the quarterly average number, F for the customer from the transaction date for the first time to the last trading date between investment in the overall number of statistics, score value Rfs for the first transaction, Rls for the last transaction score value.

$$Fq = F/(Rls - Rfs) \quad (1)$$

Because of the actual data will be the same customer at the same day of the incremental transactions, the situation will be the same day many times as a trading, the amount of investment will be graded on statistics, the aggregate amount of the Fq of 1.

In this paper, the value of the quarterly average number Fq is evaluated by using the previous score evaluation method, which is the same as Rf and Rl. As far as possible, the data of similar frequency value should be classified into a standard, the frequency of score evaluation for Fqs, specific Fq according to the score evaluation standard for processing the list below, as shown in Table III.

TABLE III. THE TABLE OF FREQUENCY DATA NORMALIZATION.

Starting Frequency	End Frequency	Benchmark Score	Customer Share (%)
0	1	1	51.308
2	3	2	45.254
4	5	3	3.037
6	7	4	0.350
8	9	5	0.035
10	11	6	0.008
12	13	7	0.003
14	15	8	0.005

(c) Standardized processing of value data

The average customer transaction amount Mq is the average value of the client's investment in financial products transactions between April 2015 and March 2017.Client average transaction amount Mq is a customer in April 2015 to March 2017 investment products trade value, the average amount of value is the amount of time interval average total amount of the investment to the value on each transaction in the each quarter. The standardized treatment of values is shown in Eq. 2, which Mq for the average transaction amount, M for the period from trading for the first time to the total amount of the cumulative value of the last transaction, Rfs for the first time trade score value, Rls for the last transaction score value.

$$Mq = M/[(Rls - Rfs + 1) \bullet F] \quad (2)$$

In this paper, the actual transaction data and product data of the bank are selected. The minimum amount of the bank's trading votes for fifty thousand yuan, so there will be no average transaction amount is less than 50000.

Value of standardized treatment in this article also uses the score evaluation method, the method of average transaction rate statistics, and the similar customer purchasing behavior as together, the score evaluation of the value of Mqs. As shown in Table IV.

TABLE IV. THE TABLE OF VALUE DATA NORMALIZATION

Initial amount	End of The Amount	Benchmark score	Customer Share (%)
50000.0	80000.0	1	19.066
80000.0	120000.0	2	15.623
120000.0	210000.0	3	19.015
210000.0	310000.0	4	12.604
310000.0	452000.0	5	10.665
452000.0	720000.0	6	10.223
720000.0	1380000.0	7	7.805
1380000.0	414660000.0	8	4.999

C. System Construction and Application of Customer Segmentation Model

The construction and application of customer segmentation model are mainly divided into two main components:

The first part is the data analysis of customer segmentation. System using historical customer purchasing behavior information for data analysis, and then the system through the pretreatment of RFM analysis score customer's transaction database, Then, the system uses the value score of the three attributes of RFM analysis to express the customer, and finally the customer cluster analysis is conducted.

The second part is the model application in the financial product recommendation guidance system. According to the customer clustering information analyzed in the first part, the customer is analyzed by the system to obtain the customer classification information. And combined with the customer's risk tolerance, the marketing personnel to carry on the sale of financial products recommendation guidance.

(1) Construction of RFM Customer Segmentation Model

Results the collection point diagram is shown in Fig. 1.

As can be seen from Fig. 1, when K=9, the calculation results fluctuate, but after the fluctuation, it tends to be stable, so we choose 8 as the value of K.

The specific process of the algorithm is shown in Table V.

In the process of k-means clustering, the selection of the initial clustering center point of step 3 will result from the different influence of the center point.In the k-means of Spark MLlib, there is a mode "Kmeansǁ" which can be selected for the initial cluster center, follow the basic principles of K-means++ algorithm of this model, the initial clustering center of the distance between each other should be as far as possible,.You also can intuitively understand that this is going to be better for the K initial cluster centers.

TABLE V. THE TABLE OF ALGORITHM STEPS.

Algorithm Step	Description
1	Read the RFM analysis score data set W.
2	Determine and select K values for clustering.
3	Initial K cluster center point selection.
4	Calculate the Euclidean distance between the data concentration points and K clusters, and select the nearest central point to divide into the cluster.
5	Calculate the mean value of the coordinates of all points in each cluster as the new center.
6	Whether the algorithm satisfies the convergence condition S, if satisfied to step 8, does not satisfy the convergence condition to step 7.
7	Determine whether the algorithm reaches the maximum iteration threshold, if it reaches the eighth step, and does not reach the step 4, 5 and 6 of the cycle.
8	Output K cluster corresponding result set.

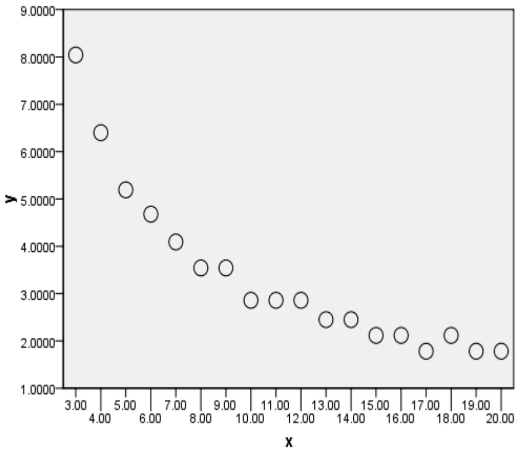

Fig. 1. The Scatter plot of the K selected results.

The basic steps are as follows:

1) Randomly select a point from the data set X as the first starting point and define X as a sample of X set.

2) Calculate the shortest distance D(x) between each x and the cluster center.

3) Calculate the probability of each x being selected as the next class cluster center

$$P(x) = \frac{D(x)^2}{\sum_{x \in X} D(x)^2};$$

Choose the largest x of P(x) as the next cluster center.

(2) Repeat the above two steps until the end of the K initial center selection.

In the process of k-means clustering, the convergence condition S is satisfied in step 6, which is usually used to determine whether the discriminant condition of the convergence of K-means algorithm is the squared error and S, and its definition is as shown in Eq. 3. Where K represents K cluster centers, wi represents the focal point of the first i cluster center, and Edist represents Euclidean distance.

$$SSE = \sum_{i=1}^{K} \sum_{x \in W_i} Edist(w_i, x)^2 \quad . \tag{3}$$

In the process of k-means clustering, the fifth step needs calculate average value of all points in each cluster, the calculation steps are as follows.

It can be seen from Eq. 3 that the optimal center of mass can be obtained when the error sum of squares is the lowest. So the formula is minimized, and the derivative of S with respect to S is equal to zero, as shown below:

$$\begin{aligned} \frac{\partial}{\partial_{W_k}} SSE &= \frac{\partial}{\partial_{w_k}} \sum_{i=1}^{K} \sum_{x \in W_i} (w_i, x)^2 \\ &= \sum_{i=1}^{K} \sum_{x \in W_i} \frac{\partial}{\partial_{w_k}} (w_i, x)^2 \; . \\ &= \sum_{x \in W_k} 2(w_k - x_k) = 0 \end{aligned} \tag{4}$$

$$\begin{aligned} \sum_{x \in W_k} 2(w_k - x_k) = 0 &\Rightarrow n_k w_k = \sum_{x \in W_k} x_k \\ &\Rightarrow w_k = \frac{1}{n_k} \sum_{x \in W_k} x_k \quad . \end{aligned} \tag{5}$$

In other words, for wi, when you take the derivative, when $w_k = \frac{1}{n_k} \sum_{x \in W_k} x_k$, the SSE minimizes the SSE's optimal center of mass is the mean of each point in the cluster. Therefore, the formula can be deduced similarly. For xi, when xi=wi, SSE is the smallest, and corresponds to step 4 in the process of k-means clustering.

In conclusion, as the number of iterations increases, the SSE value decreases to a minimum value, so SSE converges. According to the clustering results, the spatial scatter diagram is drawn to show the clustering distribution as shown in Fig. 2.

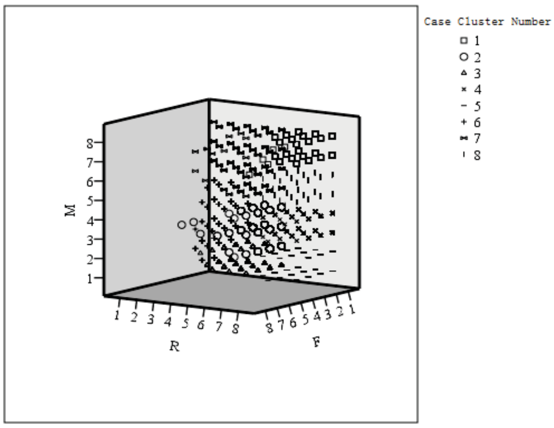

Fig. 2. The Scatter plot of clustering distribution.

(3) Model Application in Financial Product Recommendation Guidance

Customer segmentation model in financial products recommended the application of the guidance system is mainly, through the analysis of customer classification segmentation results to customers recommend products, at the same time, system can provide a guideline for marketing personnel. First, the model represents the RFM score of the customer. Next, the model cluster analysis to the customer and the results of the clustering are matched. Finally, the customer classification information is obtained. This is the specific application process of the model in financial product recommendation guidance system. The system matches the of the customer and the hierarchical information of the customer's risk tolerance, and then carries out the product recommendation according to the matching information. The specific process is shown in Fig. 3 below.

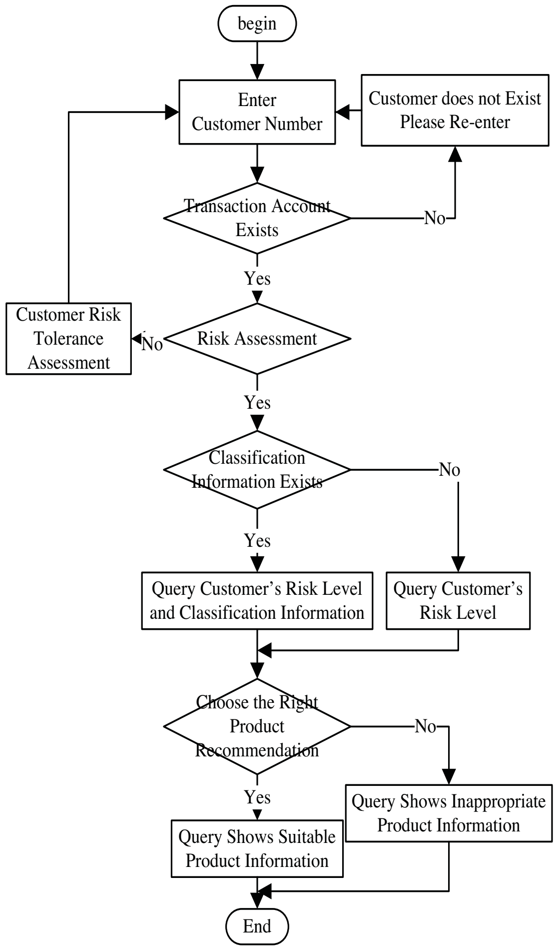

Fig. 3. The flow chart of financial management product recommendation.

Based on the clustering results, the average value of the RFM model score of the eight kinds of customers is calculated. Then the system compares the average score within the class with the average score of the whole. Finally, the customer is redefined according to the comparison result of system, and the results are shown in Table VI.

(4) Information Analysis of Financial Customer Segmentation Model

It is known from Table VI in the previous section, cluster 1 and cluster 8, two customer groups are important value customers. Their RFM model index of customer segmentation in the bank financial product recommendation guidance system is higher than the average. It shows that the recent investment of such clients is relatively close, the number of investment per quarter is higher, and the average investment amount per quarter is higher. This kind of client has a great contribution to the development of bank finance and has been formed in the bank to buy financial products consumption habits. This reflects the recognition of commercial banks' financial products.

Clustering group 7 is important call customers. In the RFM model of customer subdivision of bank financial product recommendation guidance, the freshness of investment is lower than the average value but closer to the average value. The consumption frequency and the average investment amount of each quarter are higher than the average. These clients are not very interested in recent investments, but having frequent and high investments, so they are important call customers.

TABLE VI. THE TABLE OF ANALYSIS RESULTS

Class Number	Average of Recency	Average of Frequency	Average of Monetary	Customer Number	Compare Results	Customer Level
1	7.16	1.71	7.39	3870	R↑F↑M↑	Important customer
2	4.64	1.59	3.07	4506	R↓F↑M↓	General maintenance of customer
3	3.67	1.37	1.30	6028	R↓F↓M↓	Loss of customer
4	6.89	1.55	3.45	6661	R↑F↑M↓	Potential customer
5	6.84	1.38	1.47	4882	R↑F↓M↓	General observation of customer
6	2.05	1.29	2.75	2655	R↓F↓M↓	Loss of customer
7	4.23	1.61	6.10	2774	R↓F↑M↑	Call back the customer
8	7.05	1.68	5.50	5794	R↑F↑M↑	Important customer
Total	5.60	1.53	3.67	37170		

(5) Information Analysis of Financial Customer Segmentation Model

It is known from Table VI in the previous section, cluster 1 and cluster 8, two customer groups are important value customers. Their RFM model index of customer segmentation in the bank financial product recommendation guidance system is higher than the average. It shows that the recent investment of such clients is relatively close, the number of investment per quarter is higher, and the average investment amount per quarter is higher. This kind of client has a great contribution to the development of bank finance and has been formed in the bank to buy financial products consumption habits. This reflects the recognition of commercial banks' financial products.

Clustering group 7 is important call customers. In the RFM model of customer subdivision of bank financial product recommendation guidance, the freshness of investment is lower than the average value but closer to the average value. The consumption frequency and the average investment amount of each quarter are higher than the average. These clients are not very interested in recent investments, but having frequent and high investments, so they are important call customers.

Clustering group 2 is generally retain customers. In the RFM model of customer subdivision of bank financial product recommendation guidance, the freshness of investment and average investment in the amount of each quarterly investment are lower than the average value but closer to the average value, and the investment frequency is higher . The customers have certain investment habits. Each time they can control the amount of investment and consider the investment cycle, so the customers need to maintain.

Clustering group 4 is the potential customers. In the RFM model, which is recommended by bank financial products, the investment freshness and the quarterly investment frequency are higher, but the average investment amount per quarter is lower than the average. The clients have been investing heavily in the near future, but they can control the amount of each investment. Therefore, the clients need to tap the investment potential.

Clustering group 5 is general observation customers. In the RFM model, which is recommended by bank financial products, the investment freshness is higher. But the quarterly investment frequency and the average investment amount per quarter is lower than the average. This kind of customers may be for short-term users of low investment and new customers of bank financial products which has the certain development space, so belong to need to continue to observe to the investment potential. The customers are called general value customers.

Clustering group 3 and 6 are the loss of our customers. In the RFM model, which is recommended by bank financial products, the investment freshness, the quarterly investment frequency and the average investment amount per quarter is lower than the average. So they need to be left behind.

IV. Conclusions

The system uses the improved RFM index to conduct data preprocessing to achieve the purpose of analyzing the transaction information behavior of financial clients. At the same time, the system normalizes the original data of the model, and then the system uses the k-means clustering technique commonly used in data mining to excavate the original data of the model. Then the system is analyzed according to the clustering results to obtain the customer classification information. According to the real bank financial customer transaction data, based on the closed-end investment products integrated trading behavior, and then analysis the result, will eventually analysis results and the integration of marketing strategy to establish financial RFM model of customer segmentation. On the basis of guaranteeing stability, the system improves the analysis efficiency, so that the customer classification analysis and the bank financial products recommendation and guidance management achieve effective combination. Therefore, this system has made an important contribution to the bank's further promotion of marketing service level and the exploration of customers' potential ability.

References

[1] J.Y. Bao, "On Global Financial Impact, Market Constraint and "Hostile risk" Governance of China's "One Belt And One Road" Overseas Investment", *China Soft Science*, vol. 3, pp. 1-10, 2017.

[2] X. Ruan, "Design and Implementation of Customer Relationship Management System of Guiyang Wudang Bank", *Xiamen: Xiamen University*, 2015.

[3] N.G. Mu, H. Wang, "Opportunities and Challenges in the Development of Traditional Commercial Banks under the Background of Internet Finance", *E-commerce*, vol. 2, pp. 36, 2016.

[4] M. Dietz, P. Harle, S. Khanna, et al., "Fighting for Customers: How Traditional Banks Deal With Internet Financial Challenges", *New Finance*, vol. 1, pp. 37-44, 2017.

[5] Y.G. Feng, "Empirical Analysis on the Value of the Customer's Word-of-mouth Recommendation based on Customer Lifetime Value", *Statistics and Decision-making*, vol. 12, pp. 186-188, 2016.

[6] X.Q. An, "Research on Customer Value Recognition based on RFM Model in C2C Environment", *Chengdu: Southwestern University of Finance and Economics,* 2012.

[7] L. Xiong, B. Gao, "Research on Customer Segmentation based on RFM Multi-level Customer Value Model", *Business Era*, vol. 5, pp. 55-57, 2017.

[8] H.L. Li, "Application of Data Analysis and Mining in the Refined Operation of E-commerce", *China Economic and Trade*, vol. 15, pp. 52-53, 2017.

[9] X.X.Wu, "User Segmentation of Internet Finance Platform based on RFM Improvement Model", *Beijing: Beijing Jiaotong University*, 2016.