

Research on Customer Segmentation Method of Commercial Bank Based on Data Mining

Dapeng Dong¹, Jianhua Zhang^{1,*}, and Jianwen Ye¹

¹Management Engineering Department, Zhengzhou University, Henan province, China
f_author@996934418@qq.com, t_author@1357568274@qq.com

Keywords: Data mining; Customer relationship management; K-means algorithm; Customer segmentation

Abstract: At present, in the field of commercial banks, customer relationship management (CRM) is a focus of customer-centric management of commercial banks. Who can accurately use the data mining technology to commercial bank customer information classification, mining and processing will success. And then customers should be segmented in order to timely make adjustments and improvements to adapt to market changes, and seek more business interests. To this end, this paper analyses the role of customer relationship management and data mining, as well as their current development. Through the analysis of data mining K-means algorithm and a large number of customer information summary analysis, and then much effective information is available to customers. The purpose of the subdivision is to help commercial banks carry out the implementation of CRM, which has a particularly important significance.

1. INTRODUCTION

In the mode of commercial economy, data mining technology has become an important basis for commercial banks to segment customers. China's commercial banks in the continuous improvement of their own data mining capabilities, a better understanding and grasp the dynamic changes in the market, according to customer demand changes, to quickly respond to improve the existing management model, targeted operational strategy. To help commercial banks to keep old customers, reduce customer churn, to attract new customers have a greater role, but also can help commercial banks to express their business, the maximum reduction in operating costs, access to more benefits.

2. THE APPLICATION OF DATA MINING IN CUSTOMER SEGMENTATION OF COMMERCIAL BANKS

2.1 Applied to the Basic Process of Data Mining in Commercial Bank CRM

In order to better promote the development of commercial bank customer relationship management, here the data mining in its use of the process were summarized:

(1) Determine the analysis and forecasting objectives. This is equivalent to demand analysis, mainly clear objectives. Commercial Bank CRM only to determine the target customers, in order to better to collect data, analysis and mining of hidden information.

(2) The establishment of data mining library. Commercial banks to establish a database, it is necessary to collect, categorize, process and analyze the various data of customers. Finally, the data will be processed into the establishment of the database, so that you can get a commercial bank CRM itself with a high degree of

confidentiality of a unique database.

(3) Analysis of data. with data mining technology to analyze the data in the database, according to the bank's request to extract the necessary information resources.

(4) Establishment of model. The establishment model is the result of the strategic analysis of the bank's existing business processes and operating mechanism based on the results of the data analysis, which is based on the clear development strategy of the commercial bank customer relationship management. The algorithm used in the data analysis and comparison, so as to select the appropriate model.

(5) After the model evaluation and verification model is established, it is necessary to evaluate the validity of the model. Only to analyze the data in the model can we find out whether the model can be applied in commercial bank CRM and whether it is optimal.

(6) Implementation of the model. The implementation of the model are: First, the data mining model, and from effective information to provide relevant decision-makers to help them make the appropriate decision-making. Second, save the model, you can use the model to analyze the relevant or similar problems, so you can more quickly get the results.

3. IMPROVEMENT OF K-MEANS ALGORITHM IN CUSTOMER SEGMENTATION OF COMMERCIAL BANKS

3.1 Optimization of Initial Center Selection

Traditional K-means algorithm to the customer segmentation, is randomly selected K data objects as the initial point. It is difficult to ensure the accuracy of the results, so the initial center point to optimize the selection.

Input: The set of targets containing N sample points, the number of K clusters;

Output: A set containing K initial center points;

Step 1: Find the two points with the largest distance

in the data set and use it as the initial point A_1, A_2 ;

Step 2: Select the N/K points closer to the point A_1 and divide it into the set C;

Step 3: Remove the sample points from C from the data set and find the clustering center of C.

the other data objects in the sample are placed in the set K, and finally the number of clusters K is obtained and the initial center point is found.

3.2 Detection and Processing of Noise Points and Isolated Points

Isolated points and noise points will affect the accuracy of data mining results, they are in the process of data mining, easy to deviate from most of the data results, if not care of these points as the starting point, they produce a lot of error, The exact impact of data mining analysis can be described as devastating.

Assume that the data set DS has N m-dimensional vector samples with a m-dimensional space:

$$DS = (X_1, X_2, \dots, X_n)^T \quad (1)$$

$$X_i = (X_{i1}, X_{i2}, \dots, X_{im}) \quad (2)$$

The following calculations are made in the data set DS:

$$B_{ij} = \frac{X_{ij} - \bar{X}_i}{\sigma_i} \quad (3-1), \quad B_i = \frac{1}{m} \sum_{j=1}^m B_{ij} \quad (3)$$

The results are processed and analyzed:

$$\bar{B} = \frac{1}{N} \sum_{i=1}^N B_i \quad (4)$$

And through the experimental analysis to select the appropriate threshold $\alpha > 0$, The noise point and the set of isolated points can be expressed as $DS_1 = \{B_i | B_j - \bar{B} > \alpha\}$, Then DS set expect DS_1 are normal data points.

3.3 Customer Segmentation Model Based on Improved K-means Algorithm

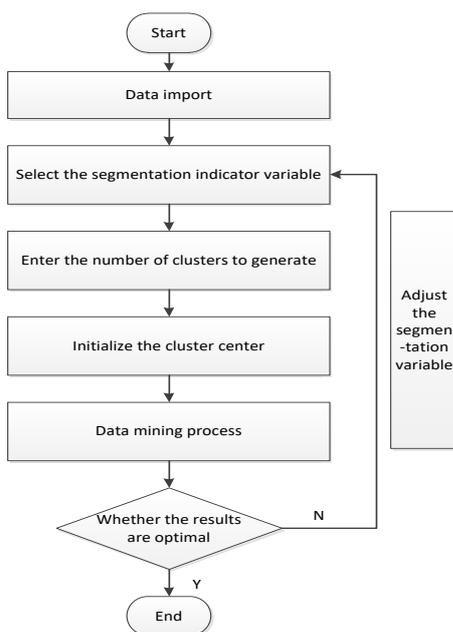


Figure 1 Customer segmentation model based on K-means algorithm.

According to the principle, process and advantages of the improved K-means algorithm used in customer relationship management of commercial banks, a customer segmentation model based on K-means algorithm is designed as shown in Figure1.

4. EMPIRICAL RESEARCH

4.1 Collection and Processing of Commercial Bank Customer Data

In this paper, Zhengzhou Everbright Bank, for example, from January 2013 to December 2015 a part of the bank branch of the customer data, randomly selected 10,000 copies of the study and analysis. Data processing is performed on the business records of the collected customers. Through the use of data cleaning, abnormal value processing, data integration and conversion, etc. do not meet the research requirements of 673 data, get accurate and reliable 9327 data. We will quantify the age, sex, marital status, education level, annual income, total deposit, total consumption amount and financial management amount of the customer in order to better analyze and deal with the data.

4.2 Empirical Results and Analysis

In this paper, SPSS software is used to do the empirical results and analysis. The data obtained from the preprocessing of the customer are imported into the Excel table. The K-means algorithm in the SPSS software is used to quickly classify the imported data. , Through the previous data analysis of the relevant data that when $K = 6,7$, the distance between the two categories is too small, the characteristics of the individual types reflected is not obvious, so the final number of clusters as 5 . After processing the data, the following results are obtained:

Table 2. Iteration History

| Iteration | Change in Cluster Centers | | | | |
|-----------|---------------------------|-------|-------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | 2.130 | 2.462 | 2.743 | 2.503 | 3.365 |
| 2 | 1.177 | 1.196 | 1.620 | 1.020 | 1.497 |
| 3 | .903 | .882 | .902 | .999 | .867 |
| 4 | .383 | .224 | .020 | .032 | .303 |
| 5 | .024 | .015 | .008 | .016 | .013 |
| 6 | .002 | .006 | .000 | .004 | .005 |
| 7 | .001 | .003 | .000 | .001 | .002 |

Table 4. Distances between Final Cluster Centers

| Cluster | 1 | 2 | 3 | 4 | 5 |
|---------|-------|-------|-------|-------|-------|
| 1 | | 4.747 | 2.842 | 4.905 | 6.435 |
| 2 | 4.747 | | 1.891 | 3.635 | 4.023 |
| 3 | 1.842 | 3.891 | | 3.511 | 6.241 |
| 4 | 3.905 | 3.635 | 3.511 | | 3.763 |
| 5 | 6.435 | 6.023 | 6.241 | 4.763 | |

Table 5. Number of Cases in each Cluster

| | | |
|---------|---|----------|
| Cluster | 1 | 107.000 |
| | 2 | 159.000 |
| | 3 | 123.000 |
| | 4 | 8548.000 |
| | 5 | 390.000 |
| Valid | | 9327.000 |
| Missing | | .000 |

4.3 Customer Segmentation Results and Countermeasures

Through the above data results and analysis, bank customers can be divided into five types: potential growth customers, general customers, intermediate customers, senior customers, VIP customers. Banks can accurately grasp a variety of customers, you can plan, targeted to take measures to get more customers. Banks can consolidate old customers and absorb more new customers through a variety of marketing campaigns.

Table 6 Characteristics and Management Strategies of Five Kinds of Customers in Commercial Banks.

| Customer name | Customer characteristics | Management strategy |
|----------------------------|--|---|
| Potential growth customers | Care about the price, like preferential, low frequency of consumption, the amount of consumption less, the greater the loss, to bring small profits bank. | To develop low-cost promotions and other activities, improve service quality, to attract such customers, and strive for conversion into the general customer. |
| General customer | Concerned about preferential activities, less consumption, less wasted, to bring smaller bank profits | Develop appropriate discount promotions, carry out more reward activities, improve customer loyalty and satisfaction, and strive to transform into intermediate or advanced mobile customers. |
| Intermediate customers | Shopping consumption frequency, spending more, like to focus on the bank a variety of new activities, is the bank's loyal customers, bring more profits. | Offer more offers and increase incentives to provide low loans and credit cards for conversion to premium and VIP customers. |
| Senior customer | Shopping frequency is small, but the amount of high consumption, like a variety of investment, to bring a lot of bank profits, is the focus of maintenance of the object | Such customers are the focus of the bank, to expand the scope of concessions and to provide personalized service to maintain good customer relations |

| | | |
|---------------|---|---|
| VIP customers | Financial and consumption of a larger amount, a huge chain of funds, the bank's various investment activities more to bring the greatest interests of banks, banks are eager to develop the focus of the maintenance object | To increase the focus on such financial investment advice, increase on-site service, networking activities, increase the rate of return on investment |
|---------------|---|---|

5. CONCLUSIONS

In the context of the business age, through the use of data mining technology, commercial banks can be collected by the customer data analysis and analysis, get more valuable information for more customers to understand the breakdown, in order to different customers for different Business strategy. K-means algorithm can help banks to better segment customer groups, identify the business value of different groups, minimize the value of low-key customers, maintain and develop more high-value customer base. So that banks can continue to gain their place in the competition, better and faster forward.

ACKNOWLEDGEMENTS

This work is supported by the Excellent Young Teacher Development Fund Project of Zhengzhou University (Foundation No.2015 SKYQ15) and the Innovation and Development Fund of Zhengzhou University School of Management Engineering (Foundation No. 20170604).

REFERENCES

- [1] Feng, F, 2012. Research of customer relationship management solutions based on data mining [M]. London: Springer London. 670-675.
- [2] Feng, Y, Na, L, 2013. Research of relationships between products and customer based on data mining technology [J]. Applied Mechanics and Materials. 2668 (401): 2372-2375.
- [3] Jie, Zhang, 2015. Improvement and Application of K-means Clustering Algorithm [J], Electronic Technique Applications. 41 (01): 125-128 + 131.
- [4] Liping, Yan, 2007. Analysis and Design of Telecom Customer Segmentation Model Based on Data Mining [J], Technology Square. (05): 89-90.
- [5] Xiaoli, Yin, 2009. Application of Data Mining Technology in Bank CRM [J], Chinese Journal of Economic Research. (20): 112-113.
- [6] Xue, Bai, 2012. Application of data mining technology in customer relationship management [J], Popular Technology. (02): 27-29.
- [7] Ying, W, 2013. Research of data mining in customer relationship management [M], Heidelberg: Springer Berlin Heidelberg. 163-170.

Table 1. Initial Cluster Centers

| Cluster | 1 | 2 | 3 | 4 | 5 |
|------------|-----------|-----------|-----------|-----------|-----------|
| AGE | 48 | 58 | 57 | 55 | 38 |
| SEX_ID | 2 | 2 | 2 | 1 | 2 |
| EDU_ID | 3 | 1 | 2 | 4 | 4 |
| MARRY_ID | 1 | 1 | 1 | 1 | 1 |
| SALARY | 1036892 | 29590 | 803395 | 3634550 | 86329 |
| AMOUNT | 244622.86 | 322101.23 | 93757.94 | 217409.28 | 302693.12 |
| COST_TOTAL | 66357.35 | 138102.73 | 225605.67 | 183494.60 | 236056.08 |
| FINANCE | 72124 | 53560 | 80386 | 54512 | 62884 |

Table 3. Final Cluster Centers

| Cluster | 1 | 2 | 3 | 4 | 5 |
|------------|-----------|-----------|-----------|-----------|-----------|
| AGE | 41 | 37 | 42 | 41 | 43 |
| SEX_ID | 1 | 1 | 2 | 1 | 1 |
| EDU_ID | 3 | 2 | 3 | 2 | 2 |
| MARRY_ID | 1 | 1 | 1 | 1 | 1 |
| SALARY | 96860 | 89509 | 111902 | 100253 | 102620 |
| AMOUNT | 239900.12 | 315461.24 | 456238.15 | 516071.67 | 443140.36 |
| COST_TOTAL | 324571.10 | 233574.04 | 278653.26 | 302672.11 | 296811.22 |
| FINANCE | 240216 | 40008 | 52262 | 128400 | 58079 |