

Customer Classification of Discrete Data Concerning Customer Assets Based on Data Mining

Zuo Lei, Guo Junfeng

Harbin Finance University, Harbin, Heilongjiang, 150030, China
523406570@qq.com

Abstract—selecting useful information under the background of big data can help enterprises to classify customers more accurately. Outlier data includes important customer information. In order to study customer classification problem based on customer asset outlier data, a customer classification model based on outlier data analysis concerning customer asset is constructed successfully. The model is based on Variables in 4 dimensions including transaction frequency, types of products or services traded, transaction amount and client age. And using clustering before classification to divide twenty-five types of outlier customer data into four categories and corresponding marketing strategies also are put forward according to different classification of outlier customer data of a company.

Keywords- Discrete data; Customer Classification; Customer Assets; Data Mining

I. INTRODUCTION

With the arrival of the era of big data, enterprises' data has formed a certain scale in the field of marketing. Its diversity, low-value density and real-time complexity are both challenges and opportunities for marketing. In marketing management classifying customer management is one of the core issues of enterprise operation. Identifying and owning excellent customers, and developing and maintaining customers in a targeted manner, not only avoids the waste of resources and higher costs caused by the decentralization of energy, but also reduces the huge risk of blind marketing. By collecting the data generated by customers and enterprises at the contact end, the effective customer classification not only filters the data interference of customers who do not have transaction relationship with enterprises in the market, but also avoids the legal risk of infringing customer privacy. In the process of customer data mining, outlier data are often encountered. They are inconsistent with the law embodied in the overall data representation level. They are free from most of the intervals and are usually considered as noise data or abnormal data to be eliminated. However, as objective data, the way of processing is obviously inappropriate.

Therefore, how to filter data from mass data and how to use data mining algorithm to complete the value "purification" of customer data and find important customers have become the urgent problems to be solved in the marketing field under the background of large data. On the basis of relevant research, a customer classification model is constructed in this paper based on customer asset outlier data analysis, and corresponding marketing strategies are put forward for different customer classification. Based on the traditional customer classification model RFM, the age dimension is added to the three dimensions including transaction frequency, types of products or services traded

and transaction amount, and the customer classification model based on customer asset outlier data analysis is constructed and customer information is deeply excavated from the perspective of outlier data, and customer classification is carried out.

II. CUSTOMER CLASSIFICATION MODEL BASED ON OUTLIER DATA ANALYSIS OF CUSTOMER ASSETS

As far as outlier data is concerned, marketing outlier data meets the characteristics of ordinary outlier data and they are free from the activity area of most common data, but they can reflect some important characteristic information of customers. For example, the information of "extremely large transaction amount" belongs to outlier information, but it reflects the strong purchasing power of customers. It is not wise to ignore such customers.

As far as customer assets are concerned, it aims to help enterprises understand customers more deeply, grasp customer needs, and intend to influence customer behavior by managing customers, so as to promote customers to make greater value contribution to enterprises. Customer classification indicators

A. Customer classification indicators

This study intends to build a customer classification model based on customer asset outlier data analysis. Four dimensions of indicators including transaction finance, types of products or services traded, transaction frequency and customer age are used to classify customers. However, the three dimensions of customer information are more operable, so the transaction amount, the types of products or services traded, and the transaction frequency are selected as variables to enter the model. Choosing customer age as the fourth dimensions because customer assets represent the sum of customer's lifetime purchase value, which can measure to be at the stage of customer's "lifetime", and then analyze the current consumption situation and consumption growth of customers. In addition, customer's age is used as the fourth customer classification indicator. It is assumed that the overall age distribution of customers presents a normal distribution. The data of four dimensions of customer information can be easily obtained through collation at the terminal contact point. Especially in the e-commerce environment, the amount of transactions, the types of products or services traded, the frequency of transactions and the age of customers can be obtained through background operations. Through the analysis of outlier data of four dimensions, the relevant characteristics of important customers are found easily.

B. Modeling Idea

Firstly, data of transaction amount, product or service type and transaction frequency of all customers are obtained, and dimensionless processing is carried out to make the data between [0, 1]. Using MATLAB programming, the reasonable range of values of three variables is determined. Among them, the satisfying conditions of a and b are $0 < a < b < 1$, $a + b = 1$. Using formula editor, outlier data is not in reasonable interval [a, b]. Taking 0.5 as the median value, the grade less than the median value is defined as "small, less, low", and the grade higher than the median value is defined as "big, many, high".

Secondly, the age as the fourth dimension is classified. Ideally, enterprises should refer to the level of local economic development at that time, conduct consumption capacity survey, and use expert analysis to reasonably judge the age structure of enterprise customers. The age structure of customers satisfies the normal distribution model, so the probability of variables between $\mu - 3\sigma$ and $\mu + 3\sigma$ is as high as 99.7%, which basically includes all customers of enterprises. On this basis, the age distribution of customers within 99.7% is divided into different age groups. For the age range from non-dimensional to [0,1], the interval of the first age is determined to be $[0, P(X < \mu - \sigma)]$, the interval of the middle age is $[P(X < \mu - \sigma), 1 - P(X < \mu - \sigma)]$, and the interval of the tail age is $[1 - P(X < \mu - \sigma), 1]$.

The customer classification in the modeling idea is as follows:

1) The determination of discrete customers

Reasonable customers and discrete customers constitute the overall customers faced by enterprises. As can be seen from Figure 1, the total customers of an enterprise are distributed within a square with a side length of 1. The length, width and height of the square represent the types of products or services traded, the amount of transactions and the frequency of transactions respectively. At the same time, each side takes 0.5 as the median value to distinguish the different attributes of "more and less", "big and small", "high and low". The shadow cube with (b-a) as the side length is the reasonable customer area, and the area outside the small solid cube is the discrete customer area.

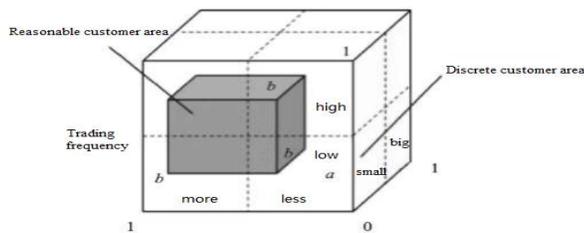


Figure 1. Enterprise customer area

The 24 types of discrete customers in Table 1 are distributed in this discrete area without considering the age dimension, and in the case of considering the age dimension, it is the area of a specific age group (first, middle and last).

2) The existence significance of reasonable customers

This study can be completed smoothly with reasonable customers. The main reasons are as follows:

First, reasonable customers exist objectively, the overall data distribution law under the condition of direct ignore will be destroyed, which will bring great trouble to the follow-up study; Secondly, reasonable customers belong to customers with good data performance in the conventional sense, which is the key part of previous research, and ignoring them directly will reduce the scientific nature of the model; Thirdly, the reasonable number of customers generally accounts for a large proportion of the total customers, which belongs to 80% of the Pareto's Law "28 Rules", the best reasonable interval cannot found with neglecting directly. According to the four dimensions of transaction amount, product or service type, transaction frequency and customer age, customers can be divided into 25 categories, and further divided into four types according to their transaction amount and age, as shown in Table 1.

Table 1 Customer classification

client	classification	client	Amount of transactions	types of products or services traded	frequency of transactions	age of customers
			Big small	more less	high low	First middle last
Client 1	First category	Customer 1	√	√	√	√
Client 2		Customer 2	√	√	√	√
Client 3		Customer 3	√	√	√	√
Client 4		Customer 4	√	√	√	√
Client 5		Customer 5	√	√	√	√
Client 6		Customer 6	√	√	√	√
Client 7		Customer 7	√	√	√	√
Client 8		Customer 8	√	√	√	√
Client 9	Second category	Customer 9	√	√	√	√
Client 10		Customer 10	√	√	√	√
Client 11		Customer 11	√	√	√	√
Client 12		Customer 12	√	√	√	√
Client 13	Third category	Customer 13	√	√	√	√
Client 14		Customer 14	√	√	√	√
Client 15		Customer 15	√	√	√	√
Client 16		Customer 16	√	√	√	√
Client 17		Customer 17	√	√	√	√
Client 18		Customer 18	√	√	√	√
Client 19		Customer 19	√	√	√	√
Client 20		Customer 20	√	√	√	√
Client 21	Fourth category	Customer 21	√	√	√	√
Client 22		Customer 22	√	√	√	√
Client 23		Customer 23	√	√	√	√
Client 24		Customer 24	√	√	√	√
Client 25		Customer 25	reasonable			

3) Management Strategy of Four Categories of Customers

According to the related research of customer relationship management, customer relationship management includes the following contents: customer relationship management is one of the contents of enterprise strategy; information technology is the method of customer relationship management; customer relationship management begins with in-depth analysis of customer behavior and characteristics, which also means interaction with customers; The purpose of customer relationship management is to achieve the balance between maximizing customer value and maximizing enterprise revenue. Different customers have different relationship values, and enterprises must focus on the most valuable customers. Therefore, based on the theory of customer relationship management, the following management strategies are proposed for four types of customers:

The first type of customer management strategy is proposed. The first category of customers mainly includes the customers whose transaction amount is large and whose age distribution is in the first and middle stages. For the first class of customers, we must adhere to the principle of "key development". The transaction amount of the first group of customers is large. Generally speaking, it is likely to be a large enterprise customer and the most important customer. Enterprises should adopt the strategy of key development to provide additional preferential and special services to customers. The first group of customers has different needs and age distribution. Enterprises can be equipped with corresponding managers to focus on supervision, understand customer's needs, actively recommend products for customers, attach importance to after-sales service, and improve customer loyalty and reputation. At the same time, It is necessary to increase the public relations of these customers, provide more thoughtful and satisfactory services for customers, handle the relationship with customers, and even form alliances.

The second type of customer management strategy is proposed. The second category of customers mainly includes the customers whose transaction amount is large and whose age distribution is at the end of the customers. The principle of "critical maintenance" can be applied to the management of the second largest category of customers. The transaction amount of the second group of customers is large, which may be an important part of the customer source of the enterprise now. But the age of the second group of customers belongs to the end. No matter how many kinds of products or services are traded and the frequency of transactions is high or low, it will be running in a downward trend for a period of time in the future. Even many existing customers enjoy the last product or service provided by the enterprise. Therefore, we should adopt the strategy of "key maintenance" for these customers, do a good job in all kinds of additional services related to products, and strive for reputation in order to bring new excellent customer assets for enterprises by customer influence.

The third type of customer management strategy is proposed. The third category of customers mainly includes

customers with small transaction amount and age distribution in the first and middle stages. The management of the third category of customers can adopt the principle of "cultivating moderation". The third kind of transaction amount is small. Generally speaking, it is not a large enterprise customer. It cannot become an important source of customers in the short term, but enterprises cannot give up on it. Among the third group of customers, they are relatively young and have certain purchasing power, so they have good plastic value. Maintain customer relationship, then tap the needs, focus on cultivating the loyalty of the third category of customers, and promote their transformation to the first category of customers. However, the third category of customer's loyalty is low. If there is no positive feedback on the cultivation of enterprises in a certain period of time, enterprises can consider not wasting too much resources and energy to develop such customers.

The fourth type of customer management strategies is proposed. The fourth category of customers mainly includes customers with small transaction amount and age distribution in the end section and other reasonable customers. The management of the fourth category of customers can adopt the principle of "moderate abandonment". The fourth kind of customers has low loyalty. According to their characteristics, they do not create objective value for enterprises at this stage, and they are expected to contribute little to their value in the future. The best strategy to manage the customers whose customer assets belong to the fourth category is to give up the principle appropriately. Instead of striving for the lifelong purchase value of the fourth category customers, we should provide them with reasonable services and use more resources for more efficient allocation so as to ensure the sustainability of enterprise marketing and reduce market risk.

III. BUILDING CUSTOMER CLASSIFICATION MODEL

According to Pareto's "28 Rules" in management, customers are divided into two groups: 80% of enterprises' income comes from 20% of customers, and 80% of customers create low profits for enterprises. Therefore, the basic classification rules for establishing the customer classification model are that the fourth largest category accounts for 80% of the total number of customers. The specific modeling process is as follows:

A. Data collection

Collate customer information forms including transaction amount, products and services traded, transaction frequency and age segments to form aggregate customer number set and single customer data set. The total number of customers set is $X = \{x_1, x_2, \dots, x_n\}$, where N is the total number of customers. The single customer data set is $x_i = [e_1, e_2, e_3, e_4]^T$, which represents the transaction amount, products and service types, transaction frequency and age segment of the first customer.

B. Dimensionless treatment

The extreme value method is applied to deal with the single customer data for infinite tempering. The formulas are as follows:

$$e_{ij} = \frac{e_{ij} - \min e_{ij}}{\max e_{ij} - \min e_{ij}}$$

C. Dimensionless treatment

Customer 21, Customer 22, Customer 23, Customer 24 and Customer 25 are defined as C_{21} , C_{22} , C_{23} , C_{24} and C_{25} in the fourth category. According to the rule of set operation, the set of the fourth largest group of customers is C. The specific set is as follows:

$$C_{21} \cup C_{22} \cup C_{23} \cup C_{24} = \begin{cases} e_{11} \leq a, e_{14} \geq 0.8413 \text{ or } e_{11} \leq 0.5, e_{14} \geq d \\ \text{or } e_{11} \leq 0.5, e_{12} \leq a, e_{14} \geq 0.8413 \text{ or } e_{11} \leq 0.5, e_{12} \leq a, e_{14} \geq 0.8413 \\ \text{or } e_{11} \leq 0.5, e_{12} \geq b, e_{14} \geq 0.8413 \text{ or } e_{11} \leq 0.5, e_{12} \geq b, e_{14} \geq 0.8413 \end{cases}$$

$$C_{25} = \{a \leq e_{ij} \leq b, 0 \leq e_{14} \leq 1, j = 1, 2, 3\}, C = C_{21} \cup C_{22} \cup C_{23} \cup C_{24} \cup C_{25}$$

In addition, using gradient descent method, initial values are given, namely $a = b = 0.5$, $C_c = 0.1587$, $d = 0.8413$ (where C is the probability of normal distribution $P(x < -1)$, $d = 1 - c$), using MATLAB programming the reasonable interval is obtained as follows:

$$\begin{cases} \min b \\ \text{or } \max a \\ \text{length}(C)/n \geq 80\% \end{cases}$$

Among them, length(C) is the dimension of C and N is the total number of customers.

D. Determining the type of customers

Firstly, the following three sets are defined.

The first kind of set, namely the set of initial judgement of scope, is:

$$\begin{aligned} A_1 &= \{i \mid e_{ij} \leq 0.5, j = 1\} \\ A_2 &= \{i \mid e_{ij} \geq 0.5, j = 1\} \\ A_3 &= \{i \mid e_{ij} \leq 0.5, j = 2\} \\ A_4 &= \{i \mid e_{ij} \geq 0.5, j = 2\} \\ A_5 &= \{i \mid e_{ij} \leq 0.5, j = 3\} \\ A_6 &= \{i \mid e_{ij} \geq 0.5, j = 3\} \end{aligned}$$

The second kind of set, namely discrete judgment set, is:

$$\begin{aligned} B_1 &= \{i \mid e_{ij} \leq a, j = 1\} \\ B_2 &= \{i \mid a \leq e_{ij} \leq b, j = 1\} \\ B_3 &= \{i \mid e_{ij} \geq b, j = 1\} \\ B_4 &= \{i \mid e_{ij} \leq a, j = 2\} \\ B_5 &= \{i \mid a \leq e_{ij} \leq b, j = 2\} \\ B_6 &= \{i \mid e_{ij} \geq b, j = 2\} \\ B_7 &= \{i \mid e_{ij} \leq a, j = 3\} \\ B_8 &= \{i \mid a \leq e_{ij} \leq b, j = 3\} \\ B_9 &= \{i \mid e_{ij} \geq b, j = 3\} \end{aligned}$$

The third kind of set, namely age subsection set is:

$$\begin{aligned} Y_1 &= \{i \mid e_{ij} \leq c, j = 4\} \\ Y_2 &= \{i \mid c \leq e_{ij} \leq d, j = 4\} \\ Y_3 &= \{i \mid e_{ij} \geq d, j = 4\} \end{aligned}$$

Secondly, 25 kinds of customers are synthetically described by three sets.

The first group of customers:

$$\begin{aligned} C_1 &= (B_3 \cup B_6 \cup B_9) \cap (A_2 \cap A_4 \cap A_6) \cap Y_1 \\ C_2 &= (B_3 \cup B_6 \cup B_7) \cap (A_2 \cap A_4 \cap A_5) \cap Y_1 \\ C_3 &= (B_3 \cup B_4 \cup B_9) \cap (A_2 \cap A_3 \cap A_6) \cap Y_1 \\ C_4 &= (B_3 \cup B_4 \cup B_7) \cap (A_2 \cap A_3 \cap A_5) \cap Y_1 \\ C_5 &= (B_3 \cup B_6 \cup B_9) \cap (A_2 \cap A_4 \cap A_6) \cap Y_2 \\ C_6 &= (B_3 \cup B_6 \cup B_7) \cap (A_2 \cap A_4 \cap A_5) \cap Y_2 \\ C_7 &= (B_3 \cup B_4 \cup B_9) \cap (A_2 \cap A_3 \cap A_6) \cap Y_2 \\ C_8 &= (B_3 \cup B_4 \cup B_7) \cap (A_2 \cap A_3 \cap A_5) \cap Y_2 \end{aligned}$$

The second group of customers:

$$\begin{aligned} C_9 &= (B_3 \cup B_6 \cup B_9) \cap (A_2 \cap A_4 \cap A_6) \cap Y_3 \\ C_{10} &= (B_3 \cup B_6 \cup B_7) \cap (A_2 \cap A_4 \cap A_5) \cap Y_3 \\ C_{11} &= (B_3 \cup B_4 \cup B_9) \cap (A_2 \cap A_3 \cap A_6) \cap Y_3 \\ C_{12} &= (B_3 \cup B_4 \cup B_7) \cap (A_2 \cap A_3 \cap A_5) \cap Y_3 \end{aligned}$$

The third group of customers:

$$\begin{aligned} C_{13} &= (B_1 \cup B_6 \cup B_9) \cap (A_1 \cap A_4 \cap A_6) \cap Y_1 \\ C_{14} &= (B_1 \cup B_6 \cup B_7) \cap (A_1 \cap A_4 \cap A_5) \cap Y_1 \\ C_{15} &= (B_1 \cup B_4 \cup B_9) \cap (A_1 \cap A_3 \cap A_6) \cap Y_1 \\ C_{16} &= (B_1 \cup B_4 \cup B_7) \cap (A_1 \cap A_3 \cap A_5) \cap Y_1 \\ C_{17} &= (B_1 \cup B_6 \cup B_9) \cap (A_1 \cap A_4 \cap A_6) \cap Y_2 \\ C_{18} &= (B_1 \cup B_6 \cup B_7) \cap (A_1 \cap A_4 \cap A_5) \cap Y_2 \\ C_{19} &= (B_1 \cup B_4 \cup B_9) \cap (A_1 \cap A_3 \cap A_6) \cap Y_2 \\ C_{20} &= (B_1 \cup B_4 \cup B_7) \cap (A_1 \cap A_3 \cap A_5) \cap Y_2 \end{aligned}$$

The fourth group of customers:

$$\begin{aligned} C_{21} &= (B_1 \cup B_6 \cup B_9) \cap (A_1 \cap A_4 \cap A_6) \cap Y_3 \\ C_{22} &= (B_1 \cup B_6 \cup B_7) \cap (A_1 \cap A_4 \cap A_5) \cap Y_3 \\ C_{23} &= (B_1 \cup B_4 \cup B_9) \cap (A_1 \cap A_3 \cap A_6) \cap Y_3 \\ C_{24} &= (B_1 \cup B_4 \cup B_7) \cap (A_1 \cap A_3 \cap A_5) \cap Y_3 \\ C_{25} &= B_2 \cap B_5 \cap B_8 \end{aligned}$$

IV. CONCLUSIONS AND SUGGESTIONS

This paper provides a method of outlier analysis aiming at providing modules to help enterprises classify customers according to customer assets, so as to identify customers with good customer assets, and then develop and maintain customers in a targeted manner, which not only avoids the waste of resources caused by decentralization, but also reduces the huge risk brought by blind marketing of enterprises. Of course, as the driving force of enterprise growth and the ultimate purchaser of services, the best strategy of enterprises is always to persevere in providing them with increasingly perfect services.

Acknowledgements

Basic research project of basic scientific research fund for Heilongjiang Province higher education, Project number is 2017-KYYWF-E0103.

REFERENCES

- [1] Guo Junfeng, Qi Jingjia, Research on the Optimization of Decision Support System Based on Data Mining Algorithms. Bulletin of Science And Technology, 2018: Vol.34, (3)192-194. (In Chinese).
- [2] Sun Xiaolin, Yao Bo, Chen Yu, Customer Classification Based on Discrete Data Mining of Customer Assets. Statistics & Information Forum, 2018: Vol.33, 114-120. (In Chinese).