Acta Technica 62 (2017), No. 6B, 913–924

Customer relationship management based on demand forecast in cloud data environment

Yanhong Zhong[1], Liuyuan Xu[1], Yongjia Lin[2]

Abstract. With the accumulation of consumption data and the development of data mining, the customer segmentation based on consumer behavior becomes more and more important. This paper puts forward a new method of customer segmentation for consumption data mining with intact process. Firstly, design the customer segmentation model with 3 kinds of types and 10 indexes, and use the factor analysis method to extract the segmentation variables, then to segment the customers by partitioning method based on division. Through the calculation and analysis of the 1 million sales data of a large-scale paper towel manufacturing enterprise, the effective customer category is obtained, which shows that this method has stronger ability of customer segmentation and interpret ability of customer behavior characteristics.

Key words. Customer segmentation, Consuming behavior, Data mining, Cluster.

1. Introduction

Expanding domestic demand, especially consumer demand, is the fundamental foothold of China's long-term steady and fast economic development. If needing to promote consumption, it is necessary to understand the consumers, and have reasonable guidance. The competitions among enterprises become more intense in today with excess production capacity and under-consumption. More and more enterprises are changing from product oriented business model to customer oriented business model. Good relationship between enterprises and customers is the key factor in the competitiveness, for this purpose, the enterprises need to solve three basic problems, namely access to customers, retain customers and maximize customer value. Due to different value of different customer, the customer segmentation can make the enter-

[1]School of Accounting, Guangdong University of Finance & Economics, Guangzhou City,Guangdong Province, 510000, China

[2]School of Entrepreneurship, Guangdong University of Finance & Economics, Guangzhou City, Guangdong Province, 510000, China

prises identify different customer groups, to provide customers with differentiated, personalized products and services, thereby increasing the customer satisfaction and loyalty.

On the other hand, people have more business data than ever before. Leading enterprises not only store these data, but also with the help of advanced data analysis system based on business intelligence (BI) to help enterprises develop and realize the enterprise strategy [?]. Customer segmentation based on data mining is both the supplement of traditional customer segmentation method and the important application of business intellectual technology. In the enterprise, the massive data analysis has been the technical bottleneck, the author has obtained the certain research achievement in the aspect of business model of business analysis system and the high performance real-time data mining [?]. On this basis, this paper proposes a new method of multi-index customer segmentation based on consumption data mining, and providing a more complete solution of customer segmentation for enterprises, to avoid the data deviation caused by human factors and improve the scientificity and veracity of customer segmentation. This paper also proved that this method has stronger ability of customer segmentation and interpret ability of customer behavior characteristic through a case study of a large scale paper towel manufacturing enterprises in China.

2. Multi-index RFM consumer segmentation based on consumption data mining

2.1. Research methodology

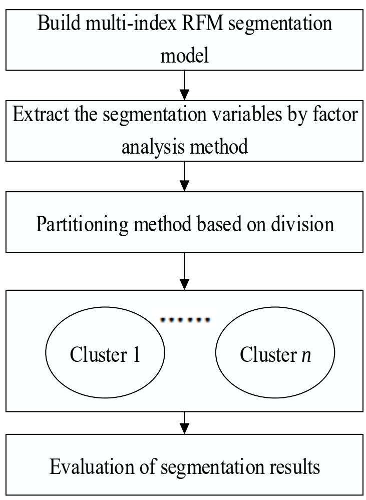


Fig. 1. Structure drawing of research methodology

This paper puts forward a new and intact method of customer segmentation for consumption data mining. Firstly, design the RFM segmentation model with

3 kinds of types and 10 indexes, and use the factor analysis method to extract the segmentation variables, then to segment the customers by partitioning method based on division. As shown in Figure 1, the detailed process can be divided into the following 4 steps: (1) building a multi-index RFM segmentation model; (2) using the factor analysis to extract the segmentation variables. The factor analysis can obtain the number of potential factors and the ability of factor to explain information, and calculate the scores of each customer on factor; (3) taking the scores of customers on factor as input data of clustering method based on division; (4) evaluating the results of customer segmentation.

2.2. Multi-index RFM segmentation model building

(1) Deficiency of traditional RFM model

The consuming behavior of customers is a very complex process, the above three indexes to some extent reflects the customer purchase and customer value, but there are still some defects. For the recent purchase indicated by R, the new customers and old customers may have the same performance, that is they simultaneously have the records of recent purchase in the period of observation, but the companies are unable to determine the new or old nature of customers in accordance with R, and provide different decisions for customers with different natures, finally resulting in the biased decisions. The frequency of consumption indicated by F can show the purchase strength of customers in a short period (such as month), but if the period of a customer is longer (such as a year), then as the number of purchase in the longer period, F will not explain the purchase rule of customer, for example, which month do customers purchase the most frequently, which month is the most weak, and which month is lower than the average purchase? Similarly, the amount of customer consumption is the same. The traditional RFM model can not reflect this information, but this information can provide the correct guidance for enterprise marketing strategy.

(2) Improved multi-index RFM model

In order to avoid these defects, based on the in-depth analysis of the traditional RFM index, this paper constructs a RFM segmentation model with 3 kinds of types of 10 indexes to reflect the customer value more comprehensively.

Table 1. Index selection

Traditional RFM index	Improved RFM index
Recent purchase R	Recent purchase Rr, farthest purchase Rf, purchase the first quartile Rq1, median purchase Rq2, purchase the third quartile Rq3
Frequency of consumption F (SUM_FREQ)	Total purchase frequency SUM_FREQ, monthly maximum purchase frequency MAX_FREQ, monthly minimum purchase frequency MIN_FREQ
Amount of consumption M (SUM_MNY_MONTHLY)	Accumulated amount of purchase SUM_MNY_MONTHLY, average amount of purchase AVG_MNY_MONTHLY

Using the multi-index to instead of the traditional consumption degree, frequency and value, mainly based on the following considerations: (1) using multiple purchase time to replace the single recently purchase, which can really reflect the purchase law on time and the overall purchase time span of customers, it is easier to distinguish between the old and new customers. (2) Using multiple frequency data to replace the overall frequency of consumption, which can better reflect the centrality of purchase, and know the month of customers purchasing most frequently, the month of weakest purchase and the relevant amount, so as to provide the basis for marketing. (3) The average amount of consumption of the users can be analyzed by the average amount. It is easy to determine the customers with a large share with the combination of the total amount of consumption.

2.3. Factor analysis method for extraction of segmentation variable

There are different opinions on the weight division problem of the three indexes of traditional RFM model. Hughes believes that the weights of RFM are consistent on the measurement of one problem, and therefore the different division has not been given. However, Stone thinks that the weight of each index is not same through the empirical analysis of credit cards, which shall be given the weight of the highest frequency, the second degree and the lowest value. Weight analysis method commonly adapted in China is the analytic hierarchy process (AHP), but compared to the determination of weight analysis method, AHP tends to formulate the program decision, and when constructing the initial characteristic matrix, the inaccuracy of weight results are caused by the great influence of human factors. When considering the index weight of multi-index RFM model, the author uses the factor analysis method to identify potential factors that affect customer segmentation, determining factors and the number of factors and the scores of customers on factors, and to determine the weight of each factor according to the degree of each factor explaining customer segmentation.

Factor analysis is a data reduction technology. By studying the internal dependency relationship among the variables, to explore the basic structure of the observation data, and represent the basic data structure by a few hypothetical variables. The hypothetical variables (i.e. factor) can reflect the main information of the original variables. According to the general steps of factor analysis, the score on each factor of the improved RFM model can be defined as:

Where, X is the improved RFM model index, FACTOR is factor, and matrix is the coefficient matrix of factor score.

2.4. Partitioning method of customer segmentation based on division

When selecting the partitioning methods, the improved RFM indexes are numeric data, not categorical attribute data, and hope that the customer types do not overlap each other, so we use the partitioning method (Partitioning Methods) based on division to segment customers. This method takes the scores of each index on factor as the judgment basis, and segment customers with the strong similarity in a category, customers with the weak similarity in different categories, to ensure the minimum distance in the same category. The most commonly used heuristic partitioning method based on division is k-means algorithm and k-medoids algorithm. K-means algorithm is simple and fast, but it is easy to be affected by noise points and isolated points, and is not suitable to find the cluster with non-convex shape. K-medoids algorithm is not to represent the cluster with the center of mass, but with a data object that is closest to the center, which eliminates the sensitivity to isolated points and has strong robustness, but has the higher cost of execution. Typical k-medoids algorithm has PAM algorithm and CLARA algorithm, etc. In order to deal with large data sets, CLARA method can be used to cluster, by improving the heuristic function, or combining with other methods, such as rough set method, genetic algorithm, Tabu algorithm and other algorithms to further enhance the clustering effect.

3. Case study

3.1. Application example

In order to illustrate the customer segmentation method of multi-index RFM model, the sales data in 1 year of current analysis point in the sales fact table of a large scale paper towel manufacturing enterprise in China are selected as the data source. The data set has a total of 1 million sales records, covering 5355 customers, and the total sales is 1,291,304,000 yuan.

After extraction, conversion and descriptive statistics on the source data, Table 2 can be obtained as follow:

Table 2. Sales data descriptive statistics

	Minimum	Maximum	Mean	Std. Deviation
Rf	115	360	234.06	48.597
Rq1	115	360	217.85	38.892
Rq2	115	359	200.63	30.287
Rq3	115	359	184.64	23.436
Rr	1	359	170.61	22.348
SUM_FREQ	1	5589	186.74	248.407
MIN_FREQ	1	1039	44.15	56.868
MAX_FREQ	1	1531	85.96	90.684
SUM_MNY_MONTHLY	7	2.E7	2.41E5	692040.20
AVG_MNY_MONTHLY	7	1.E7	8.73E4	240417.56

After KMO test on the source data, the KMO=0.7 is obtained, which is suitable for factor analysis. Through the analysis of variance decomposition factor extraction, the information utilization rate of the former three factors reaches to 83.369%, for the three common factors, analyzing in accordance with the factor loading matrix, the Table 3 is obtained as follow:

Table 3. Factor loading matrix

Index	factor1	factor2	factor3
Rf	0.799	0.374	-0.133
Rq1	0.884	0.313	-0.099
Rq2	0.942	0.208	-0.023
Rq3	0.89	0.011	0.109
Rr	0.645	-0.242	0.223
SUM_FREQ	-0.014	0.895	-0.321
MIN_FREQ	-0.347	0.704	-0.291
MAX_FREQ	-0.24	0.859	-0.316
SUM_MNY_MONTHLY	-0.086	0.651	0.714
AVG_MNY_MONTHLY	-0.188	0.566	0.773

In the analysis of factor loading matrix, the Rf, Rq1, Rq2, Rq3 and Rr have larger positive load on the first common factor1, and other indexes on the factor1 has relatively small load, the factor1can be called as the purchase time factor, which explains all information about time. Purchase time factor can be used to determine the customer is new or old, and explain the purchase vitality; on the second common factor2, the SUM_FREQ, MAX_FREQ and MIN_FREQ have large load, while other indexes only have medium positive load, meanwhile, the load of Rr on the factor is negative recently, the reason is that the larger purchase frequency is, the shorter the purchase recent time is and vice versa. This factor can be interpreted as purchase frequency factor; on the third factor3, the SUM_MNY_MONTHLY and AVG_MNY_MONTHLY have great positive load, while the rest indexes have smaller load, this factor can be interpreted as the purchase amount factor.

In the multi-index RFM model, the customer classification is determined by the comparison between principal factor mean of each customer category and total principal factor mean, the result may be equal to or greater than (expressed with $\uparrow$) and less than (expressed with $\downarrow$) two categories, so according to the number of principal factor m, the number of cluster n can be set as 2^m kinds. The number of factors in this case is 3, and the customer can be divided into 8 types, as shown in Table 4.

After that, using the factor score formula to calculate the scores of customers on factor, then using the partitioning method based on division to build the customer segmentation model, Table 5 shows the relevant statistical data of the customer after segmentation according to the factor scores, which lists the number of customers in each type of customer segmentation, and the scores on each factor.

Table 4. Clustering type

Cluster	Type
000	factor1↓factor2↓factor3↓
001	factor1↓factor2↓factor3↑
010	factor1↓factor2↑factor3↓
011	factor1↓factor2↑factor3↑
100	factor1↑factor2↓factor3↓
101	factor1↑factor2↓factor3↑
110	factor1↑factor2↑factor3↓
111	factor1↑factor2↑factor3↑

Table 5. Multi-index RFM segmentation results

Cluster	factor1	factor2	factor3	Quantity	Ratio	Type
1	-1.0371	-4.214	25.308	1	0.02%	001(1)
2	-1.0854	-5.657	39.602	1	0.02%	001(1)
3	0.6702	2.852	0.0846	227	4.24%	111(7)
4	-0.5124	-0.497	-0.128	2570	47.99%	000(0)
5	0.6924	0.25	9.9656	13	0.24%	111(7)
6	-0.3458	0.625	-0.024	1372	25.62%	010(2)
7	-0.1142	0.548	2.9686	106	1.98%	011(3)
8	1.5441	-0.262	-0.157	1065	19.89%	100(4)

3.2. Result analysis

According to the results analysis in above table, it can be seen that there are 6 types of customer groups divided into the following: 001 (cluster 1, 2), 111 (cluster 3, 5), 000 (cluster 4), 010 (cluster 6), 011 (cluster 7) and 100 (cluster 8).

On the basis of the above classification results, we conduct statistics for purchase amount of customers in each type and the proportion of the number of people, to analyze the consumption characteristics of customers in each type.

Type 001: important big customers. The customers of this type are the latest new customers, the number of purchases is less, but the purchase amount is very large. In the two enterprises, the number of customer of type 001 accounts for about 0.04% of the total, but the amount contribution of customers of both reaches 3%, customers of this type are the main force of enterprise.

Type 111: important old customers. The customers of this type are the old customers, their last purchase time is farther, purchase frequency and purchase amount are higher than the average value, accounting for 20% of enterprise sales contribution, and the proportion of number of people is relatively small, the average contribution is high, the customers of this type are the cornerstone of enterprise, so enterprises shall focus on the development of relationships with these customers, to

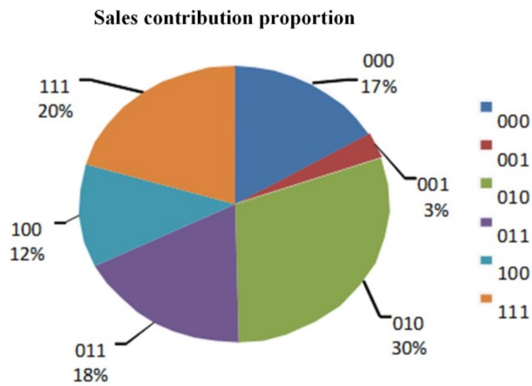


Fig. 2. Sales contribution proportion diagram

promote satisfaction and loyalty of them.

Type 000: new customers with general value. The customers of this type show that the new entry, low purchase frequency and small purchase amount. The number of customer of this type accounts for 47.99% of the total, which is the largest customer group of enterprise, but the sales contribution of this group just account for 17%, the average sales contribution of customers is low, so the enterprises can appropriately adjust the resources allocation, and transfer the resources put on the customers of this type.

Type 010: important developing customers. The customers of this type show that new entry and the higher purchase frequency, but low purchase amount. The number of customers of this type accounts for 25.62%, at the same time, the sales contribution of customers group of this type accounts for 30%, it is the customer group with highest contribution.

Type 011: important new customers. The customers of this type are performed with new entry, relatively high purchase frequency and purchase amount, although comparing with customer of type 001, the sales contribution of each customer is low, but the customers of this type occupy a certain proportion, meanwhile the sales contribution to the enterprises reaches 18%, they are the cornerstone of the enterprises, so the enterprise can increase the resource investment in the customer of this type, so as to increase profits.

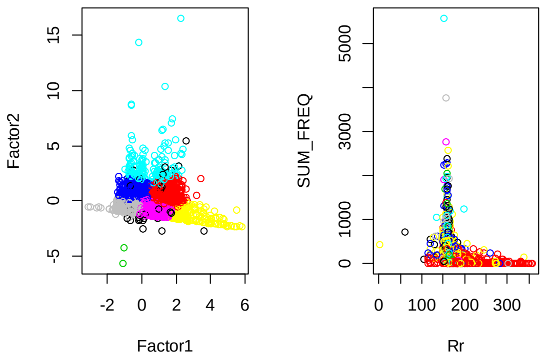
Type 100: old customers with general value. The customers of this type show that having entered for a long time, but the purchase frequency and purchase amount are low. They account for 19.89% of the total number of customers, and the sales contribution is 12%, the average contribution of customers is low. At the same time, as a result of the farther purchase time, it is likely to have been lost.

4. Segmentation effect contrast of traditional and multi-index RFM model

Continue to compare with the traditional RFM and the multi-index segmentation model using the data of case, the segmentation results of two methods are as shown in Table 6 and Figure 3:

Table 6. Customer segmentation category contrast

	RFM segmentation category	011	011	011	001	001	011	011	100	Total number of customers
Multi-index segmentation category	Cluster based on division	1	2	3	4	5	6	7	8	
001	1				1					1
001	2					1				1
111	3	97		33					97	227
000	4	114		1					2455	2570
111	5		6				4	3		13
010	6	338		13					1021	1372
011	7	12	12	82						106
100	8	112		11					942	1065
Total number of customers		673	18	140	1	1	4	3	4515	5355



(a) Segmentation results of multi index model
(b) Segmentation results of traditional RFM model

Fig. 3. Customer segmentation effect contrast

According to the analysis of above diagram, it can be seen that comparing with the traditional RFM segmentation model, the multi-index segmentation model has the following characteristics:

(1) Effectiveness. For example, the results of important big customers of type 001 are consistent by the two methods. The results show that, as in the traditional RFM, the multi-index RFM model is effective in segmentation, and can effectively segment the customers with obvious characteristics.

(2) Deepen the segmentation category. The traditional RFM model segments customers into 3 types (011, 001, 100), but the new method of multi-index customer segmentation segments customers into 6 types (001, 111, 000, 010, 011, 100), making the segmentation more detailed, to ensure that the big behavior difference between the different segments, and small behavior difference within same segmentation. In addition, after improving the RFM model segmentation, only 106 customers still are type 011 in the 838 customers of 011 in the traditional category, the remaining 351 customers are divided into type 010, 115 customers are type 000, 143 customers are type 111, 123 customers are type 100. This is more conducive to clear understanding of customer behavior for enterprises.

(3) More significance of segmentation. The segmentation of traditional RFM model presents the bipolar development, such as type 100 has 4515 customers, type 011 has 828 customers, but type 001 only has 2 customers. This result does not have the significance of segmentation. However, the number of various types of customers in the multi-index model is reasonable, which is more suitable for the formulation of marketing strategy.

To sum up, through the comparison of the two models, it shows that the new method of multi-index customer segmentation is superior to the traditional RFM model in identifying customer behavior characteristics and segmentation. Therefore, using multi-index customer segmentation model to effectively segment customers, promoting the enterprises to apply different customer development, maintain and retain strategy for different types of customers, and to better keep the customer loyalty and enhance the enterprise value.

5. Conclusion

The behavior segmentation method based on consumer data mining predicts the future behavior of customers in terms of the past and present behavior of customers. This paper designs the RFM segmentation model with 3 kinds of types and 10 indexes, and use the factor analysis method to extract the segmentation variables, then to segment the customers by partitioning method based on division, providing the enterprises with a more complete solution for customer segmentation, to avoid the data deviation caused by human factor, and improve the scientificity and accuracy of customer segmentation. The instance shows that the multi-index customer segmentation model based on consumer data mining has more effective detection rate than the traditional RFM model, and stronger explanatory ability of customer behavior characteristic. Today has entered the era of big data, which not only brings technological challenges for data mining, but also analyzes the customer consumption behavior for enterprises, and maximizes the customer value to play a greater role. Customer value analysis based on consumer data mining in big data environment will be an important research direction.

References

[1] Cao J, Wu Y, Li M: (2013) *Energy Efficient Allocation of Virtual Machines in Cloud Computing Environments Based on Demand Forecast*[J]. Journal of Chinese Computer Systems, 7296(4):137-151.

[2] Raza M Q, Khosravi A: (2015) *A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings*[J]. Renewable & Sustainable Energy Reviews, 50:1352–1372.

[3] Mohaupt M, Hilbert A: (2013) *Integration of Information Systems in Cloud Computing for Establishing a Long-term Profitable Customer Portfolio*[J]. Iaeng International Journal of Computer Science, 40(2):124-133.

[4] Niu D: (2014) *Demand Forecast, Resource Allocation and Pricing for Multimedia Delivery from the Cloud*[J]. Proceedings Iccasp, , 234(10):587-598.

[5] Lam H Y, Ho G T S, Wu C H, et al.: (2014) *Customer relationship mining system for effective strategies formulation*[J]. Industrial Management & Data Systems, 114(5):711 - 733.

[6] Song, Hua-Ming, Yang, et al.: (2011) *Optimal Ordering Policy and Coordination Mechanism of a Supply Chain with Controllable Lead-Time-Dependent Demand Forecast*[J]. Mathematical Problems in Engineering, 2011(1024-123X):701-719.

[7] Duan Z P, Qian X M, Duo Y Q, et al.: (2014) *Research on Demand Forecast for Emergency Resources Based on CBR and Grey Relational Analysis*[J]. Advanced Materials Research, 962-965:3050-3053.

[8] Fiig T, Härdling R, Pölt S, et al.: (2014) *Demand forecasting and measuring forecast accuracy in general fare structures*[J]. Journal of Revenue and Pricing Management, 13(6):413-439.

[9] Kaya M, Özalp Özer: (2009) *Quality risk in outsourcing: Noncontractible product quality and private quality cost information*[J]. Social Science Electronic Publishing, 56(7):669–685.

[10] Niu D, Xu H, Li B, et al.: (2011) *Risk management for video-on-demand servers leveraging demand forecast*[C]// International Conference on Multimedea 2011, Scottsdale, Az, Usa, November 28 - December. DBLP, 2011:1229-1232.

[11] Morenochaparro C, Salcedolagos J, Trujillo E R, et al.: (2011) *A Method for the Monthly Electricity Demand Forecasting in Colombia based on Wavelet Analysis and a Nonlinear Autoregressive Model*[J]. Ingeniería, 16(2):págs. 94-106.

[12] Peng R, Xiong L, Yang Z: (2012) *Exploring Tourist Adoption of Tourism Mobile Payment: An Empirical Analysis*[J]. Journal of Theoretical & Applied Electronic Commerce Research, 7(1):21-33.

[13] Antoniolli P D: (2016) *Information Technology Framework for Pharmaceutical Supply Chain Demand Management: a Brazilian Case Study*[J]. BBR - Brazilian Business Review, 2016, 13.

[14] Meng F, Xu C, Tang L: (2013) *Sphericity Evaluation Using Minimum Zone Solution*[J]. International Journal of Advancements in Computing Technology, 5(7):40-47.

[15] Steel S J, Louw N: (2002) *An integrated approach to outlier identification and variable selection in discriminant analysis*[J]. South African Statistical Journal, 36(2):129-151.

Received May 7, 2017