Research on Customer Segmentation Method for Multi-value-Chain Collaboration

Lei Duan[1,2], Wen Bo[1,3], Qing Wen[4], Shan Ren[5], and Changyou Zhang[1](✉)

[1] Laboratory of Parallel Software and Computational Science, Institute of Software, Chinese Academy of Sciences, Beijing, People's Republic of China
changyou@iscas.ac.cn

[2] School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, People's Republic of China

[3] School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi, People's Republic of China

[4] Chengdu Guolong Information Engineering Co., Ltd., Chengdu, Sichuan, People's Republic of China

[5] School of Information Science and Technology, Southwest Jiaotong University, Chengdu, Sichuan, People's Republic of China

Abstract. For multi-value-chain collaborative business in automotive industry, the value-based customers segmenting has become an important method to improve the synergy efficiency of automobiles. In order to accurately discover the value of potential customers, this paper proposes a customer segmentation method for multi-value-chain collaboration. Firstly, we screened evaluation index with high degree of customer value relevance, and establish a value-based customer data representation model according to customer's information we collected on the collaborative marketing platform of the automobile marketing value chain; Then, according to the distribution characteristics of the customer information, we used improved initial centroid selection method for k-means algorithm to establish customer segmentation method. Finally, based on the customer data accumulated on the car collaborative business platform, design an experiment to verify the accuracy of customer segmentation. The result of experiment shows that the customer segmentation method effectively reduces the computational complexity. This method can guide the designing of multi-value chain coordination mechanism for customer segmentation and create more value of both automobile production value chain and sales value chain.

Keywords: Multi-value-chain · Customer segmentation · Customer representation model · Cluster analysis · K-means algorithm

Y. Sun et al. (Eds.): ChineseCSCW 2019, CCIS 1042, pp. 197–211, 2019.
https://doi.org/10.1007/978-981-15-1377-0_15

1 Introduction

The automobile services include three stages: pre-sales service, in-sales service and after-sales service. Pre-sales service as the link between marketing and sales is very important for the entire automobile service chain. Automobile pre-sales service is a series works that stimulate customer's purchase desire before they contact the product. The main purpose of pre-sales is to accurately grasp and satisfy customer's actual needs. The international competition in automobile consumer market makes the traditional pre-sales service method hard to acquire the subtle needs of users. It needs to be improved by modern information technologies. The competition between automobile companies has gradually become the competition between industry chains. Pre-sales service is at the forefront of the entire sales value chain, and is also the window between automotive products and customer service brands. It is the closing point of the automotive product life cycle. Therefore, automotive pre-sales services are critical to understand user's ability and build a user-centric product service strategy. Moving the focus to customer service is helpful to keep existing customers and discover potential customers. Provide different customers with personalized services can create value for customers and maximize enterprises' profits. It becomes an effective method to enhance the competitiveness of enterprises, and enhance the value creation ability of the automobile sales value chain. In 1950s, American scholar Wendell Smith proposed the theory of customer segmentation. Customer segmentation refers to the company formulating precise strategic strategies based on its own business operation model, classifying customers according to various factors such as customers' natural characteristics, value, purchasing behavior, needs and preferences in certain marketing, and providing targeted products, services and marketing models to customers within the limited resources [1–3]. Customer segmentation in automobile pre-sales services can help companies develop precise strategic strategies based on their business models. On the one hand, customers are classified according to their natural characteristics, value, purchasing behavior, needs and preferences to provide targeted products, services and marketing models. On the other hand, it needs to identify potential customers' acquirements and provide them with accurate and satisfactory services.

2 Related Works

2.1 Related Research on Customer Segmentation

In the research of customer segmentation, scholars from various countries have proposed a variety of methods [4]. Commonly, the enterprise selects the necessary customer segmentation variables and indicator systems from many segmentation variables according to the background of the enterprise customer base. The normal customer segmentation methods include: simple statistics-based customer segmentation, behavior-based customer segmentation, value-based customer segmentation, data mining-based customer segmentation, etc. [5].

The simple statistics-based customer segmentation method is based on demographic characteristics such as age, gender, and region, or based on social characteristics such as occupation, education, marital status, and occupation. Behavior-based customer segmentation methods classified customers primarily based on customer's behavior characteristics. Value-based customer segmentation is the way that companies use for themselves, and in most cases, it used with behavior-based customer segmentation. The data mining-based customer segmentation extracts the segmentation variables to describe customers from their basic attributes and purchase behavior. Using the relevant data mining technology to establish the customer segmentation model [6], which divides customers into different categories [7].

2.2 Cluster Analysis

The traditional customer classification methods either divide customers into high, medium and low categories according to the purchase price, or divide into new customers and old customers based on purchase date. Both two methods are rough and subjective, hard to accurately predict consumption behavior of customer bases. Clustering is unsupervised learning, no label in advance, compute cluster based on the similarity between data. The objects in same class have a relatively high similarity, while the objects in different classes have lower similarity [8]. The cluster analysis can be divided into many kinds [9, 10].

Partitioning based-clustering. For a given data set and the value of categories k, the data set can be divided into k different categories according to a similarity measure, ensuring that each category has one data item at least and each data item can only belong to one category [11]. Hierarchical-based clustering. This is also a common method in clustering, which is to hierarchically decompose a given data set and form each data object into a clustering tree [12]. Density-based clustering. The main idea of this method is to divide data objects into similar categories according to the density of the data set. If the density in the "neighborhood" exceeds the threshold, the merge will continue [13]. Model-based clustering. This method can be used to locate each category using the density function of the data space, or can determine the number of clusters from standard statistics [14]. Network-based clustering. This type of method is applied to any attribute data, and the method is fast in processing, can identify categories of any shape, and input fewer parameters, less affected by isolated points.

According to the data and practicality provided by the enterprise, this paper chooses the k-means algorithm based on partition [15], which is a relatively classic unsupervised clustering algorithm, also called distance-based clustering algorithm. The method is easy to implement, and the algorithm has good efficiency and scalability on large data sets. Therefore, it has been widely used in customer segmentation [16–18].

2.3 The Main Idea of K-means

The k-means algorithm uses Euclidean distance [19] to measure the similarity, the closer the two objects are, the higher the similarity. The basic idea is to first determine the number k of clusters and select k initial cluster centers from the data sets. Then calculate the distance from the dataset to the k initial cluster centers, and merge the dataset to the cluster where is closer. Re-adjust the cluster center of each cluster until the cluster centers are not changing, indicating that the clustering is ended and get k classes [20].

This algorithm works best if the objects of the dataset is dense and the difference between the classes is significantly larger. Even facing the relatively large dataset, the algorithm is relatively scalable and efficient, and the complexity of the algorithm is O (nKI). Calculating the new cluster center is the same time complexity as calculate the clustering criterion function value, and the time complexity required is O(nd).

3 The Customer Segmentation Model

3.1 Construction of Customer Segmentation Model

In this paper, the customer value segmentation model is established for automobile sales depend on the improved k-means algorithm. The constructed indicators of this model are basic attributes and consumption behavior of the customer. The basic attributes include customers' gender, age, annual income, education level, etc. Consumer behavior is the automobile model they purchased and the spent. Taking each customer's basic attributes and consumption behavior attributes as a dimension, the basic data about the customer becomes a multi-dimensional space, and then using the clustering algorithm to build the customer segmentation model.

Represent customer information as a customer space $\{C1, C2, \ldots, Cn\}$, where each C represents a customer's attribute dimension. A concept class $\{cl_{s1}, cl_{s2}, \ldots, cl_{si}\}$ can be determined according to the customer space, where $cl_{si}(1 \ll i < k)$ represents a concept cluster, a group of customers [22], such as "High-income people". If the customer belongs to a concept class, then all attribute dimensions in the customer space can determine which cluster it belongs to in the concept class. This process completes the customer segmentation and finally performs a functional analysis of the results of the classification.

According to the model established above, the customer segmentation mainly uses the clustering algorithm to analyze the historical customer data of the enterprise, and constructs a model that can predict the future category of the customer and applies it to the enterprise marketing. The managers of the enterprise analyze the results of the subdivision to formulate targeted marketing strategies, provide suggestions for hierarchical services for different segments of the customer segmentation, and dig out the maximum value of customers.

3.2 The Improvement of Customer Segmentation Algorithm

The improvement of this paper is based on the theory that the farthest point is the least likely to be assigned to the same cluster [23]. This method can effectively prevent the objective function from falling into local optimum, also avoids that the selected points are too close, causing multiple selected cluster centers in the same class. The process of selecting the initial clustering center for the k-means algorithm by this improved method is as follows: first select the two points p and q which are the farthest distance as the first two initial cluster centers, and record them as $V_1 = o_p$, $V_2 = o_q$, $d_1 = d_{pq}$.

Then, all the remaining points are classified into V1 and V2 as the center [24], and two classes are recorded as d_{21} and d_{22}. For any point i $\in \{1, 2, \ldots, n\}$ (excluding p, q), if $|o_i - V_1| < |o_i - V_2|$, then o_i is attributed to the V_1 class, otherwise it belongs to V_2. Then calculate the distance from the data in d_{21} to V_1 and the distance from the data in d_{22} to V_2:

$$\begin{cases} d_{21} = max\{|o_i - V_1|,\, o_i \in D_{21}\} \\ d_{22} = max\{|o_i - V_2|,\, o_i \in D_{22}\} \end{cases} \tag{1}$$

$d_2 = \max\{d_{21}, d_{22}\}$ then the corresponding data is recorded as V_3 as the third initial cluster center. And so on, it is executed repeatedly until k initial cluster center points are selected.

3.3 The Effect of Improved Clustering Algorithm

To verify the feasibility and effectiveness of the initial clustering center selection, in this paper we selected three sample sets which are Balance scale, Iris and Wine datasets from UCI [25] for comparative analysis. Table 1 indicates the size, category numbers, and dimensions of the three sample sets, while Table 2 lists each cluster's data number of three sample sets.

Table 1. The basic attribute of sample set.

Sample set name	Sample size	Number of clusters	Dimension
Balance-scale	625	3	4
Iris	150	3	4
Wine	178	3	13

Table 2. The cluster number of sample set.

Clustering category	Balance-scale	Iris	Wine
First cluster	49	50	59
Second cluster	288	50	71
Third cluster	288	50	48
Total	625	150	178

Using three sample sets to finish experiments, the comparison is mainly based on the stability and accuracy of the clustering results. The accuracy of the clustering result is calculated as follows:

$$P = \frac{n}{N} * 100\% \tag{2}$$

P is accuracy, n is the correct number of classification, and N is the total number of samples. The clustering results of two methods are shown in Table 3.

Table 3. Experimental simulation results.

Number of clusters	Experimental data					
	Balance-scale data		Iris data		Wine data	
	Initial center	Accuracy	Initial center	Accuracy	Initial center	Accuracy
1	2585864	0.4276	1915119	0.5612	5016174	0.5213
2	230416549	0.6140	14514142	0.6183	8950173	0.5725
3	10630509	0.6178	13982139	0.7010	15017857	0.6500
4	375138437	0.6193	1410963	0.5633	991453	0.6123
5	91389179	0.4789	10587145	0.7988	16715170	0.5217
6	1651889	0.5988	11940139	0.6944	104140154	0.6210
7	509145387	0.5990	1723147	0.5074	167106140	0.7011
8	1642530	0.4856	1409645	0.5744	1096184	0.7011
9	220517439	0.4956	7913250	0.6700	40121125	0.5997
10	360116509	0.4891	1146769	0.6597	9079127	0.6590
Average accuracy		0.5358		0.6359		0.6160
New method accuracy		0.6048		0.8900		0.6730

As we can see from the above table, the method of randomly selecting the initial cluster center makes the clustering result unstable and a relatively low accuracy. The method we adopted to select the initial clustering center in this paper is better than the original method and is relatively stable. It can be applied to handle a large amount of data generated in enterprise production, and can produce relatively good customer segmentation results in the automotive industry.

4 Verification of Automotive Industry Collaboration Platform

4.1 The Choice of Customer Segmentation Variables

For different industries, there are many attributes can describe the characteristics and differences of customers. So it results different customer segmentation variables and numbers. According to the data of automotive industry and internal business of the enterprise, this paper considered various factors of customers. Based on customer's basic data provided by the enterprise and the data quantifiable principle, selected the segmentation variables from two aspects: the customer's basic attributes and the car purchase behavior. Customer's consumption amount and purchase frequency can be used to measure the current value to the enterprise, and basic attributes such as gender, income, age and education can be used to measure its future value to the enterprise. In customer's profile, the information includes the vehicle purchase records and service personnel information. Before clustering, we need to find some basic attributes of customers according to the kind of business and the quality of historical data. The basic attributes include name, age, region, gender, income, car brand, car model, color, purchase amount, vehicle use, etc. In the first step, we need to extract customer-related data and integrated into a customer information table as the data source for establishing customer segments.

For raw data, first filter out useless attributes such as name, customer number, and so on. Obtain a data source table that can be analyzed, and select the following attributes as the cluster sample attributes as shown in Table 4.

Table 4. Cluster sample attributes.

Attribute name	Attribute value range	Type of data
Gender	{Male, Female}	Class attribute
Age	{0–100}	Numerical type
Income (ten thousand yuan)	{ ≥ 0 }	Numerical type
Address	{Provinces, Municipalities, Autonomous regions}	Class attribute
Transaction amount (ten thousand yuan)	{Integer greater than 0}	Numerical type
Number of transactions	{Integer greater than 0}	Numerical type
Brand	{A B C D E}	Class attribute
Color	{Company's existing model color}	Class attribute
Model	{The models produced by the company}	Class attribute
Education	{Primary school, Junior high school, High school, Bachelor, Master's degree, Doctor}	Class attribute
Whether to buy again	{0, 1}	Numerical type

4.2 The Discretization and Normalization of Data

According to the information in the enterprise's existing database, we extracted 11 variables from the database. These data not only have numeric attributes but also other types of attributes. However, the k-means algorithm is a distance-based clustering algorithm and cannot handle non-numeric data. Therefore, in practical applications, the customer data needs to be extracted, cleaned, converted, and also needs to be discretized [26], which means some non-numeric-type attributes are converted into integers by using some data encoding method. Some of the attributes are handled as follows in Table 5.

Data discretization does not consider the importance and relevance of the relevant attributes, so the discretized data needs to be normalized. At the same time, it can reduce the number of iterations and improve the convergence speed. The normalization is to map the data value of a certain attribute to a specific range, and eliminate the deviation of the clustering result due to the difference size of numerical value. This method is mainly used in neural networks and distance-based classification and cluster mining. K-means algorithm we used is a distance-based algorithm. Therefore, the normalization process can eliminate the unfair clustering effect caused by different value ranges of each attribute.

Table 5. Cluster sample attributes.

Indicator name	Indicator value
Gender	1: Male 2: Female
Age	1: [0–30] 2: [30–40] 3: [40–50] 4: [50–65]
Education	1: Primary school and below 2: Junior high school 3: High school 4: Associate and above
Model	1: A 2: B 3: C 4: D 5: D
Address	0001: Chengdu 0101: Mianyang 0102: Deyang
……	……

We used a common method for data normalization which is Z-score normalization [27]. The method is to normalize the data set's mean to 0 and the variance to 1. The processed data conforms to the standard normal distribution. The normalization method is shown in following equation:

$$x' = \frac{x - \mu}{\sigma} \tag{3}$$

The x' represents the value of x after normalization, μ represents the mean, and σ represents the standard deviation.

4.3 The Result of Customer Segmentation

According to the past business experience of AA companies on the automobile industry chain collaboration platform, the company's customers are divided into 4 categories, and the value of K equals 4. We summarized the results for each cluster, as shown in Table 6. The area counted in this table have relatively good sales. The top three digits of the car model number in the enterprise are used to classify the product categories as A, B, C, D, and E, and the color of the model is relatively fixed.

Table 6. Cluster sample attributes.

Cluster number	1	2	3	4
Number of clients	7894	15675	10396	5851
Client ratio	19.83%	39.39%	26.11%	14.70%

4.4 Results Analysis

According to the different subdivision results, the categories are divided as Table 7:

Table 7. Analysis of clustering results.

Cluster number	1	2	3	4
Number of clients	7894	15675	10396	5851
Client ratio	19.83%	39.39%	26.11%	14.70%
Average consumption	4.6w	8.8w	6.8w	11.6w
Average age	37	42	50	40
Male	86.86%	87.02%	86.28%	82.53%
Female	13.14%	12.98%	13.72%	17.47%
Main area	Sichuan 14.4% Shandong 7.9% Henan 7.5%	Sichuan 19.5% Chongqing 9.6% Guizhou 8.9%	Sichuan 15.3% Hebei 6.8% Henan 9%	Sichuan 17.8% Chongqing 13% Henan 6.4%
Main product category	B 25.67% C 20.98%	B 35.78% D 12.45%	A 35.78% B 19.24%	C 29.47% E 16.29%
Color	Plain white 32.0%	Pearl White 40.3%	Pearl White 26.1%	Pearl White 40.1%
Best seller	SQJ6451B	SQJ6460C	F16-T01	B60X-T02

Evaluate the different characteristics of each category, we can realize that the second and third types of customers accounted for 65.5% of the total customers, and the transaction amount accounted for 78.69% of the total transaction, which indicates that these two types are the core customers of the enterprise. From the average age of customers, most of them from 37 to 50 years old, and male customers' accounts for a large proportion. The color of the automobile is mainly white. From the aspect of customer's area distribution, customers are concentrated in Sichuan, Chongqing, Henan, Shandong and other regions. From the product categories that customers mainly purchase, A, B, and C are the main products of the company, while B is the best-selling product. It can be seen from the Table 7 that the customer group of the company is monotonous, and the scope of the covered consumer groups is very narrow. Most customers are men, lack of female customers, and the colors are mainly dominated by white. Based on the above statistics, each type of customer has the characteristics of its consumption behavior. Different consumption behaviors can be viewed through the customer information attribute in the clustering result. Relevant personnel in the enterprise can formulate marketing policies according to the attributes of customers, and maximize the value created by the customers. In addition, the manufacturer can analyze customer's purchase characteristics, understand their target customer's needs, and seize the market to provide relevant service.

5 Supporting Service Quality Based on Customer Segmentation

After completing the segmentation of customers, the statistical analysis method is used to analyze the data between different customer groups. Enterprises need to use limited resources, making corresponding marketing strategies and implementation, to achieve the actual value of customer segmentation, improve the service quality of enterprises. According to customer's commonality and characteristics, sales personnel made corresponding marketing plans for different customer groups to achieve differentiated marketing services. Combined the potential customer management module to predict the purchase will of potential customers, optimize the original sales business process and create more value.

5.1 Differentiate Marketing Strategies

Customers can be divided into four types through customer segmentation: core customers, high-end customers, normal customers, and potential customers.

The core customers made greatest contribution to the company and is the main source of enterprise's profits. In normal times, it is necessary to improve their service quality and let them experience a sense of superiority. The measures include set up a VIP area and point strategy, or give a certain repair discount, etc. To ensure that core customers can be retained for a long term, and enhanced customer loyalty. The high-end customers have relatively high purchasing power and most of them are male. According to that, the enterprise can provide them certain products or maintenance services to keep them and create greater value for the company. The normal customer's transaction amount is low, but their amount is large. They have the characteristics of dissociation, always make comparison of corresponding products from multiple enterprises, and pay attention to the sales activities. Therefore, enterprise can develop some promotion strategies to increase the consumption propensity of normal customers. The potential customers are also a huge market. According to the car model recommended by the system, set a tracking plan to provide them relatively correct requirements in an accurate and timely manner, making them become the real customer of enterprise.

5.2 Evaluation and Analysis of Service Quality

The customer segmentation can improve service quality, differentiated services can be provided for different types of customers to improve their satisfaction. Develop more potential customers while retaining existing customers, thereby maximizing the profits and enhancing corporate competitiveness. To evaluate the effect of customer segmentation in pre-sales service, this paper proposes a pre-sales service evaluation system based on the enterprise service capability model. The service evaluation system includes service operation and service performance, as shown in Fig. 1.

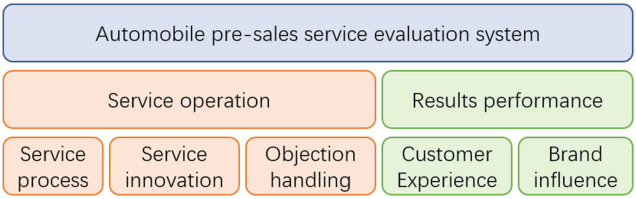

Fig. 1. Automobile pre-sales service evaluation system.

The service operation factors are used to evaluate the specific service operation of the enterprise, including customer evaluation of service processes, service innovations, and the disputes in services. Service performance factors are the qualitative or quantitative evaluation used to evaluate service outcomes, including customer experience and brand impact. The system default weight of each evaluation factor item given by the enterprise in Table 8 (the weight of each evaluation element can be adjusted according to the needs of different enterprises):

Table 8. Service quality evaluation element weight.

Service process	Service innovation	Objection handling	Customer experience	Brand influence
2	2	1	4	1

Based on the automobile pre-sales service evaluation system, we randomly selected 2,000 customers from AA companies to finish service quality assessment studies. After segmenting 2000 customers, the result shows that core customers accounted for about 43%, high-end customers accounted for about 27%, normal customers accounted for about 18%, and potential customers accounted for about 12%. Dividing 2,000 customers into two parts (maintaining the same customer's proportion of four types), 1000 of them provide unified pre-sales services, and another 1000 customers provide differentiated services according to their type. In the group which provide unified pre-sales service, we collect and summarize points of the service quality scored by customers, and shown in Fig. 2.

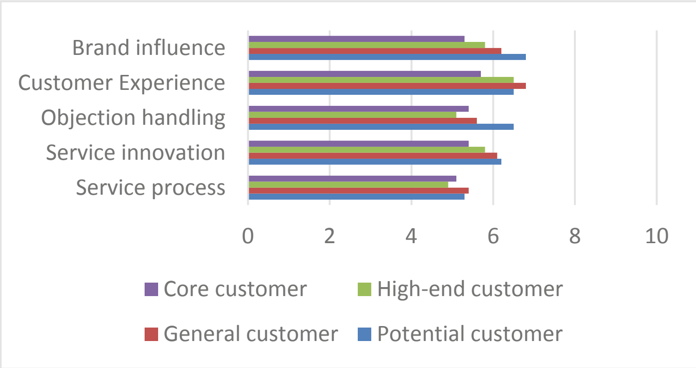

Fig. 2. Traditional service quality score table.

The result is, 235 customers have purchased the car of this company, and effective customers account for 23.5%. In the customer group with differentiated marketing strategy, provide targeted customer service for different customer types, and summarize points of the service quality scored by customers, and shown in Fig. 3.

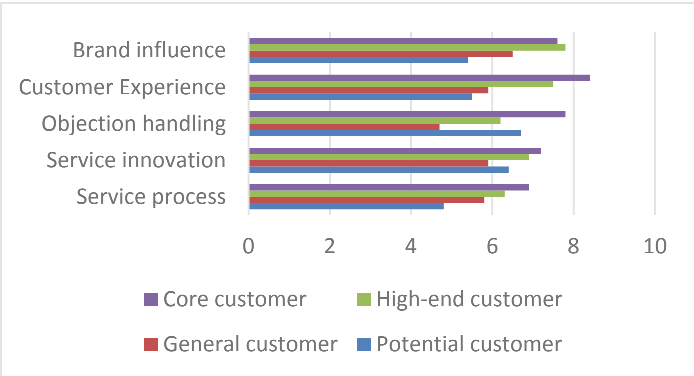

Fig. 3. Differentiated service quality score table.

The result is, 371 customers have purchased the car of this company, and effective customers account for 37.1%.

5.3 Collaboration of Multiple Value Chain

Customers evaluate the service quality of multiple sale providers, and multiple enterprises achieve the collaboration of marketing value chain through the evaluation system. From the above part, we can know that the service quality score table is based on the service quality assessment and the conversion rate of effective customer. After customer segmentation, using differentiated marketing strategies, the service satisfaction of potential customers and general customers has been slightly improved, but high-end customers and core customers have been greatly improved. It effectively aroused more purchase desires of high-end and core customers, which stimulated healthy competition among enterprises and prompted the quality of service. At the same time, it has effectively improved the satisfaction of customers, converted more potential customers into effective customers, and brought more customer resources to entire enterprises group. The customer segmentation reduced the waste of human resources, improves the efficiency of enterprises, and promoted the collaboration efficiency of the automobile multi-marketing value chain.

There are two main tasks of the manufacturing factory, one is the development of new cars, and the other is automobiles production. The development of new cars must first go through market research, figure the model and market goals, and then start the designing. Finally, the quality inspection begins after a series of adjustments, and after the quality inspection is passed, new car's data is entered into the database and arranged by the company's business department for production. The business department issues the production task to the production workshop, and the purchasing department submits the applications to purchase the materials and parts. After the production workshop produces a car, it has to pass the quality inspection, and then register and put into storage. When receiving the shipping order, the logistics department is responsible for transporting the car out of the warehouse. The above process constitutes the production value chain of the manufacturing factory. The sales department provides a series of services such as registration information, product introduction, test drive, and handles the delivery procedures after customer placed the order.

To improve the operation process of internal value chain, the company strengthens the links between various parts and establishes the industrial collaboration value chain as we shown in Fig. 4.

Based on large amount of customer data, dealers finished customer segmentation to integrate customer's feedback and market information. Sharing customer's feedback data to the new vehicle develop department of the manufacturer, it will provide a reference for the future development, which will help the manufacturer to design a car that better meets the market demand and bring a greater benefit to the manufacturers and dealers. At the same time, the dealers sharing the data with business management department of the manufacturer. Through the research on customer's feedback and market requirement, the business department adjusts the promotion and inventory of each type of car in the next quarter. Reduce the possibility of inventory backlog of vehicles while ensuring sufficient sales volume. The business department will share the main production type of next quarter and part of publicly available sales plan to the dealer, and guarantee to provide sufficient cars to dealers. When introducing the product to customers, the dealers will focus on the main production type. This measure will reduce the decrease of customer satisfaction caused by the lack of stock and the effect of brand reputation. Through the interaction process, the collaborative relationship between the manufacturing value chain and the marketing value chain is established to achieve a win-win goal.

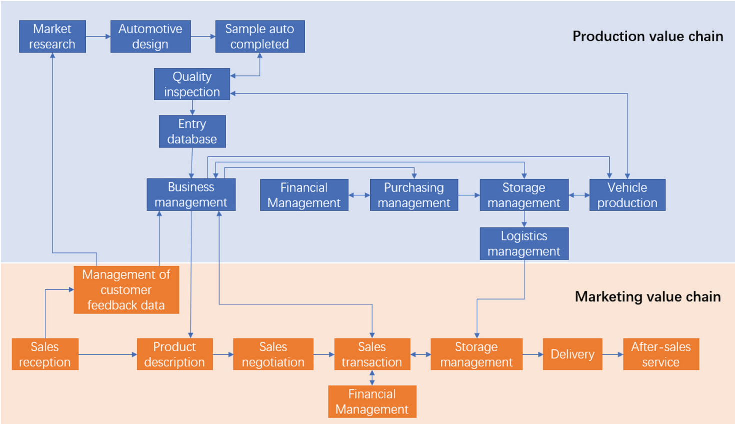

Fig. 4. Value chain collaboration.

6 Conclusion and Prospect

Verification experiment shows that cluster analysis using K-means algorithm can achieve a well effect of customer segmentation in automotive multiple value chain collaborative services. The results of customer segmentation and the consumption behavioral characteristics of each type of customer will effectively improve the collaborative efficiency between the marketing value chain and the manufacturing value chain, provide individualized and accurate services for customers. This method has a strong practical guiding significance. The successful application of cluster analysis will promote the development of customer relationship management in automotive service industry and optimize the collaboration between multiple value chains.

Acknowledgment. The authors would like to thank the anonymous referees for their valuable comments and helpful suggestions. This paper is supported by The National Key Research and Development Program of China (2017YFB1400902).

References

1. Parsell, R.D., Wang, J., Kapoor, C.: Customer Segmentation, WO/2014/099928 (2014)
2. Wang, J., Yin, L., Xu, J.: Research and Application of LNG Customer Segmentation Based on Cluster Analysis. Zhonghai Petroleum Gas and Electricity Group Co., Ltd. (2019)
3. Chen, Z.-Y., Fan, Z.-P., Sun, M.: Multi-kernel support tensor machine for classification with multitype multiway data and an application to cross-selling recommendations. Eur. J. Oper. Res. (11), 34–37 (2016)
4. Ou, J., Cao, X., Zhang, J., Ding, C.: Research on power customer segmentation based on hybrid neural network. Comput. Digit. Eng. (2019)
5. Zhang, H.: Research and Application of Clustering ensemble Algorithm in Customer Segmentation. Anhui University (2016)

6. Zhang, L., Ma, Y.: Research on aviation customer churn and segmentation based on data mining technology and implementation of R language program. Math. Pract. Theory (2019)
7. Fan, X.: Research on Customer Segmentation and Marketing Strategy of H Bank Based on Cluster Analysis. East China University of Science and Technology (2016)
8. Zhang, Y., Zhou, Y.: A review of cluster algorithms. J. Comput. Appl. (2019)
9. Li, Z., Zhang, Y.: Analysis and evaluation of clustering analysis algorithm. Electron. Technol. Softw. Eng. (2019)
10. Wang, X., Wang, H., Wang, J.: Comparison of clustering methods in data mining. Comput. Technol. Dev. **16**(10), 20–22 (2006)
11. Lai, X., Gong, X., Han, L.: Genetic algorithm based K-medoids clustering within MapReduce framework. Comput. Sci. **44**(3), 23–26 (2017)
12. Madan, S., Dana, K.J.: Modified balanced iterative reducing and clustering using hierarchies (m-BIRCH) for visual clustering. Formal Pattern Anal. Appl. 1–18 (2015)
13. Li, W.: The Study of Clustering Algorithm based on Density. Hunan University (2010)
14. Sharma, V.P., Sharma, S, Kumar, S., et al.: Cob web and dry bubble diseases in Lentinula edodes cultivation-a new report. International Society for Mushroom Science (2016)
15. Guo, D.: Research on Algorithms of Recommendation in Personalized Recommendation Systems. Beijing Jiaotong University (2017)
16. Wang, J.: The Application Research of Cluster Analysis in Caifupai Customer Segmentation. University of Electronic Science and Technology of China (2016)
17. He, L.: Customer segmentation of China telecom. Southwest Jiaotong University (2016)
18. Huang, Y.: Analysis of airline customer value based on k-means clustering method. Commun. World, 303–305 (2018)
19. Ma, C., Wu, T., Duan, M.: Clustering algorithm based on membership degree of K-nearest neighbor. Comput. Eng. Appl. **52**(10), 55–58 (2016)
20. Mokdad, F., Haddad, B.: Improved infrared precipitation estimation approaches based on k-means clustering: application to North Algeria using MSG-SEVIRI satellite data. Adv. Space Res. (2017)
21. Viegas, J.L., Vieira, S.M., Melício, R., Mendes, V.M.F., Sousa, J.M.C.: Classification of new electricity customers based on surveys and smart metering data. Energy (09), 123–125 (2016)
22. Lu, H.: Study on value-added service for power customer segmentation. Guangdong University of Technology (2016)
23. Huang, M., Lin, H.: A new method of k-means clustering algorithm with events based on variable time granularity. In: Web in formation Systems and Applications Conference. IEEE (2017)
24. Fan, Z., Sun, Y.: Clustering of college students based on improved K-means algorithm. In: Computer Symposium. IEEE (2017)
25. UCI Data Set. [EB/OL]. http://archive.ics.uci.edu/ml/
26. Wu, C., Dong, S., Li, C.: Research on the discretization algorithm of association classification based on Gaussian mixture model. J. Chin. Comput. Syst. (2018)
27. Liu, J., Zhang, K., Wang, G.: Comparative study on data standardization methods in comprehensive evaluation. Digit. Technol. Appl. (2018)