

Customer Classification of a LTL Freight Enterprise Based on K-means Clustering Algorithm

¹Qiang Liang and ²Wang Zeng,

^{1,2}Logistics Engineering, Beijing Wuzi University, Beijing, China

Abstract: With the continuous expansion of the LTL freight market and the intensified market competition, the fierce market competition has transformed the LTL freight market into a customer-oriented market. This has made companies pay more and more attention to customer relationship management, and customer segmentation is part of corporate customer relationship management. Therefore, this paper selects the customer data of a LTL freight company, uses the K-means clustering algorithm to subdivide the company's customers into three categories: basic customers, key customers, and VIP customers with different types and characteristics. Propose the use of differentiated marketing strategies, formulate different marketing strategies for each type of customer, and allocate corporate resources to various types of customers reasonably, so as to fully meet customer needs while improving the core competitiveness of the company.

Keywords: LTL freight, k-means, customers

I. INTRODUCTION

Through the analysis of the LTL freight industry, it can be seen that the LTL freight market is growing. The growth is accompanied by problems. Many LTL freight companies have begun to take the path of transformation, and more diversified choices have emerged. Some people are looking for new business models. , Some people do the whole vehicle business; there are also looking for capital support and so on. The combination with big data and other technologies is also one of the diversified development paths of LTL freight companies. This article will use data analysis technology to classify corporate customers through a LTL freight company and improve the company's customer development strategy in a targeted manner.

II. DATA AND METHODS

A. Enterprise customer instance data

The data in this article comes from the waybill information of a certain month in 2020 of a less-than-carload freight logistics company. There are more than 1,000 pieces of waybill information. This article selects 120 pieces of information as examples for cluster analysis. The waybill information includes customer account number, waybill number, business volume, volume, billable weight, main originating station, main terminal station, starting city, destination city, service type, settlement method, revenue station, income, recent cooperation Information such as month, date of recent cooperation, revenue generating station, and loss situation. Because of the large amount of data, the clustering results obtained after clustering are too complicated, and it is not easy to draw the law. Therefore, the five variables of representative freight order number, business volume, volume, billing weight, and income are selected as cluster analysis. variable.

B. k-means clustering algorithm

1. Basic principles of k-means algorithm

The k-means clustering algorithm is an iterative solution clustering analysis algorithm and an unsupervised learning method. The K-means algorithm takes k as the input parameter and divides the set of n objects into k clusters, so that the similarity within the result cluster is high, but the similarity between the clusters is low. The similarity of a cluster is a measure of the mean value of the objects in the cluster, which can be regarded as the center of mass or center of gravity of the cluster.

Table 1: Output results of k-means clustering algorithm

Group		Group1	Group2	Group3
Number of samples		90	24	6
Proportion of samples		75%	20%	5%
Cluster center	Waybill volume	188.55556	666.58333	1183.3333
	business volume	727.57778	1438.8333	2440.5
	Cargo volume	20.166667	43.833333	85
	billable weight	117.05556	354.91667	762.66667
	revenue	305.5	657.79167	1194.5

The K-means algorithm uses distance as the evaluation index of similarity, and considers that the closer the distance between two objects, the greater the similarity. The algorithm considers that clusters are composed of objects close to each other, so it takes as the final goal to obtain compact and independent clusters.

2. k-means Algorithm Steps

(1) Select the number of clusters k; (2) Generate k clusters arbitrarily, and then determine the cluster centers, or directly generate k centers; (3) Determine the cluster center point for each point; (4) Recalculate its new cluster center; (5) Repeat the above steps until the convergence requirement is met, usually the convergence requirement is to determine the cluster center point and not change it.

III. CUSTOMER CLASSIFICATION BASED ON K-MEANS CLUSTERING ALGORITHM

A. k-means cluster analysis steps

Use Ranaconda software to perform clustering analysis of K-means algorithm to realize the subdivision of the customer group of this example.

(1) Standardized data

Since the five variables selected in the information table vary greatly, each variable in the data collection must be standardized to a variable with a mean value of 0 and a

standard deviation of 1 through the scale() function.

(2) Determine the number of classes

This is a key issue in cluster analysis. The "elbow method" is used to determine the value of k. A graph with k on the abscissa and SSE (sum of squared errors) on the ordinate shows an arc-shaped drop. Curve. It has an inflection point, or there is a position where the downward trend is no longer obvious, the abscissa K value corresponding to this position is

the K we chose, so the number of clusters is determined to be 3.

(3) Obtain a clustering solution

By using the anaconda software to call the k-means clustering algorithm, input the K value, and output the clustering results. Extract the subgroups, and obtain the clustering scheme as shown in Table 1.

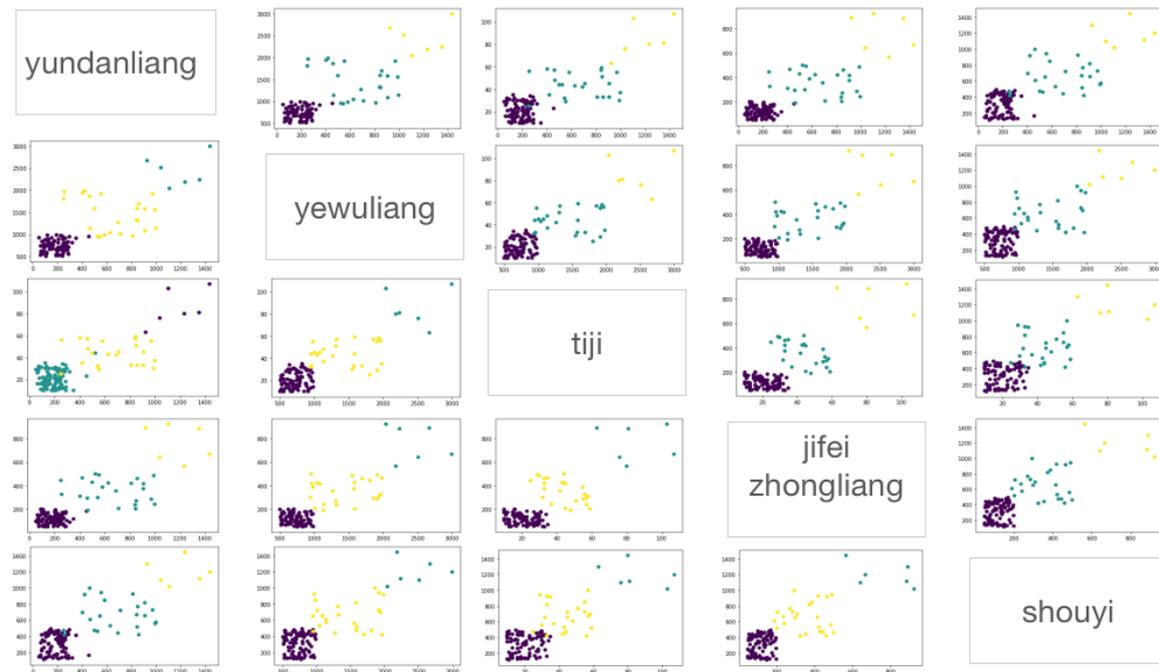


Figure 1: Scatter plot of k-means clustering results

B. k-means clustering result analysis

Through the clustering results in Table 1, it is not difficult to see that the customers of the LTL freight company are clustered into three categories, and the proportion of each category is different. Group 1 has the highest proportion, reaching 75%, but in terms of the number of shipments, business volume, volume, and billable weight, they are the least among the three groups, and the income is naturally the smallest, indicating that this group is Basic customers with the largest proportion are also so-called ordinary customers. Although the indicators of sub-group 2 are at an intermediate level, accounting for 20%, the business volume and revenue are considerable, and they are a type of group that enterprises must seize, and they are defined as key customers here. The three groups account for the smallest proportion, only 5%, but the number of waybills, business volume, cargo volume, billable weight, and revenue are the highest, which is defined as VIP customers.

By drawing a visual clustering sample scatter diagram as shown in Figure 1, the clustering characteristics and rules can be analyzed again. For example, the billable weight is the X-axis and the business volume is the Y-axis. At this time, the customer group has been divided into three categories with purple, blue, and yellow marks, which means that the three types of customers are clustered under the combination of these two levels. The number of customers marked in purple is large, the billing weight is between 50 and 300, which is the bottom end, and the corresponding business volume is between 500 and 1200, which is also the lowest. Therefore, it can be

inferred that it is a basic customer. In the same way, the number of customers marked with blue color is less than that of red mark but greater than that of yellow color mark. The billable weight is between 300 and 600, and the business volume is between 1200 and 2100. It is deduced that they are key customers. However, the number of blue label customers is the smallest, the billable weight is between 600 and 1000, and the business volume is between 2100 and 3000, so they are VIP customers. The mutual comparison and correlation of other variables are also the same. Logistics companies can observe the classification of the overall customer base under the five variables based on this, and pay attention to the influence of outliers (that is, abnormal points that deviate from the conventional segment) on the whole to find out The reason for their outlier.

IV. DIFFERENTIATED MARKETING STRATEGY

For different types of customers, the LTL freight company should appropriately adjust its corporate strategy, distinguish between services, prices, customer relationship maintenance, feedback and response strategies, and make optimal use of the company's limited resources to create for the company Greater value. Customers below provide differentiated services for various types of customers through six types of product strategy, pricing strategy, distribution channel strategy, service marketing strategy, marketing management process strategy, and communication strategy. The content of the differentiated service strategy of this LTL freight company is shown in Table 2.

Table 2: The content of the differentiated service strategy of the LTL freight company

Customer type				
The core content of differentiated services	Customer level	Basic customer	key customers	VIP customers
	Direction of development	Use the smallest investment to keep and cultivate customer relationships as much as possible, so that they can have further relationship transformation	Improve customer satisfaction, cultivate customer relationships, and make them hopefully transformed into major customers	Long-term and stable strategic partnership to achieve increased customer loyalty
	service method	Standardized service	Standardized service	Door-to-door service, customized service, innovative service, additional service and special service
	Price Strategy	Strictly follow industry regulations and standardized pricing within the company	Average level, real-time price change can be used in special circumstances	Flexible price positioning, pricing based on customer needs and corporate benefits
	Information exchange	less	frequently	Very frequently
	Staffing	No special person	Middle management	Senior management or professional docking team
	Feedback	Depends on the situation	Pay attention and respond quickly	The most value and the fastest response
	Credit rating	Lower	standard	high

CONCLUSION

The competitive landscape of China's freight market has developed to a stage where it has never been open to excessive competition. The transfer of railways to railways will inevitably lead to part of the road transport vehicles that occupy most of the freight market today. Idle vehicles will once again enter the freight market, and the competition in the freight market will become increasingly fierce. Coupled with the continuous acceleration of industry integration, the living space of small and medium-sized LTL freight companies will be squeezed once again. Therefore, the most important thing for the existing LTL freight companies is to survive and then seek development. Technological progress can often quickly promote the development of enterprises. As a new technology, big data analysis technology will be combined with LTL freight industry to improve efficiency and improve efficiency. Contribute to the internal reform of the enterprise, the development of external markets, and the improvement of core competitiveness.

Using cluster analysis to classify the existing customers of an enterprise and provide differentiated services is one of the methods to provide the core competitiveness of an enterprise. Clustering algorithm has been widely used in market, behavioral science, biology, medicine and other researches. This article uses python language to use K-means clustering algorithm to classify the customer groups of a LTL freight company. The research results show that:

(1) The clustering effect of K-means is good, which can make the customer homogeneity between the clusters and the customer heterogeneity between the categories, which is operability and reference;

(2) In this example, logistics customers are clustered into basic customers, key customers, and VIP customers. The distribution characteristics of different groups are analyzed through probability density function diagrams and sample scatter point

distribution diagrams, and the differentiation of logistics customers is proposed. Marketing plan

(3) The python language has fast clustering efficiency and accurate results.

References

- [1] K-LRFMD: Method of Customer Value Segmentation in Shared Transportation Filed Based on Improved K-means Algorithm[J].Hong Li, Xiaosheng Yang, Yao Xia, Lujie Zheng, Guoqing Yang,Pan Lv. Journal of Physics: Conference Series.2018(1)
- [2] Statistics-based CRM approach via time series segmenting RFM on large scale data[J]. Meina Song, Xuejun Zhao, Haihong E, Zhonghong Ou. Knowledge-Based Systems. 2017
- [3] Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis[J].Peiman Alipour Sarvari,Alp Ustundag, Hidayet Takci.Kybernetes.2016(7)
- [4] K-LRFMD: Method of Customer Value Segmentation in Shared Transportation Filed Based on Improved K-means Algorithm[J].Hong Li,Xiaosheng Yang,Yao Xia, Lujie Zheng, Guoqing Yang,Pan Lv.Journal of Physics: Conference Series.2018(1)
- [5] Shared-nearest-neighbor-based clustering by fast search and find of density peaks[J].Rui Liu,Hong Wang, Xiaomei Yu. Information Sciences.2018
- [6] Improved K-means algorithm based on density Canopy[J].Geng Zhang, Chengchang Zhang,Huayu Zhang. Knowledge-Based Systems.2018