

Research on Selection and Configuration Optimization of Intelligent Express Cabinets Based on K-Means Cluster Analysis*

Bo Wei

Hunan Modern Logistics College
Changsha, China

Abstract—The size and configuration ratio of intelligent express cabinets directly affect the use turnover efficiency and customer satisfaction of express cabinets. Taking the cabinet ratio of express cabinets as the research direction, this article investigates the delivery situation of courier in different regions and uses SPSS software to perform K-means clustering to obtain the size ratio of cabinets for express cabinets in different regions to meet the various demands of customers.

Keywords—express cabinets; size ratio; K-means clustering

I. INTRODUCTION

The development of the Internet has promoted the development of Internet-based e-commerce. Therefore, not only a variety of platform shopping sites have emerged, many traditional retail industries have also begun to develop in the direction of e-commerce. As a new type of consumption mode, online shopping is gradually accepted by the general public. More customers are inclined to the convenience of online shopping and gradually accept this shopping method.

The development of e-commerce has provided strong support for China's logistics and express delivery industry. As a new mode of express delivery at the end of the express delivery industry, express cabinets facilitate the delivery and collection of express parcels. Intelligent express cabinet is also called express delivery box and express self-collection cabinet. It is a service device that users can operate on their own. It is generally located in a public place with a large number of people, and its basic functions include delivery and receipt of parcels. The courier can self-deliver as needed, and the recipient can flexibly choose the time and place for collection within 24 hours, thereby solving the problem of inconvenience for immediate pickup by most people. In addition, its additional functions can provide convenience for a wider range of users. There are many types of express cabinets currently on the market. However, due to the lack of clear guidance, the service capabilities of various operators are different, which makes the customer experience quite different and seriously restricts the long-term development of express cabinets. The size and quantity of existing express cabinets are relatively fixed, but the specifications of various products are quite

*Fund: Project supported by the China Society of Logistics and China Federation of Logistics & Purchasing in 2018 (Project Number: 2018CSLKT3-228).

different. There are also huge differences in the amount of express delivery in different places such as communities and office buildings, which makes the delivery space of express cabinets very limited. The size and delivery ratio of the cabinet of the express cabinets directly affect the convenience of customers' picking up and the turnover efficiency of use.

II. THE IMPACT OF THE SIZE AND QUANTITY RATIO OF INTELLIGENT EXPRESS CABINETS ON CUSTOMER SATISFACTION AND USE TURNOVER RATE

Although there are many operators of express cabinets, and many types of express cabinets are also placed in densely populated places such as districts and office buildings, at present, the size of express cabinets on the market is basically divided into three specifications: large, middle and small. The number varies from dozens. And all kinds of products have different specifications and shapes. The characteristics of the crowd in different places such as communities and office buildings are different, and there are also huge differences in the specifications and quantity of express parcels, which makes the space for express cabinets very limited. Some businesses and users are in a wait-and-see state, and to a large extent they are concerned about the size of the cabinets, which affects the future development of the cabinets. The existing express cabinet is composed of a main cabinet and a subsidiary cabinet. A standard set of express cabinets has about 50 to 100 cabinets. The cabinet sizes of express cabinets are basically divided into small, middle and large. Small-sized cabinets can hold items that do not exceed 100mm in height, medium-sized cabinets can hold items that do not exceed 400-500mm in height, and the largest-sized cabinets can hold items that do not exceed 1m³. There are also large differences in the proportion of various storage compartments.

In the process of using express cabinets, one of the most dissatisfied factors for customers is the size limitation of express cabinets. They believe that the size ratio of existing express cabinets is relatively limited and it is difficult to meet the complexity of online shopping products. And different types of regions, consumer groups and consumption characteristics are different. If it is possible to design a suitable ratio of express cabinets according to different regional characteristics, it will increase customer satisfaction to a certain extent, and will also greatly improve the turnover

efficiency of express cabinets during use, thereby reducing the vacancy rate of some dimensions of the express cabinet.

III. OPTIMIZATION OF SIZE AND QUANTITY RATIO OF INTELLIGENT EXPRESS CABINETS BASED ON K-MEANS CLUSTER ANALYSIS

A. Questionnaire Survey and Data Statistics

1) *Design of questionnaire survey:* This study uses the form of a questionnaire, collects data through the questionnaire, and uses different analysis methods for analysis. Relatively speaking, the courier uses the express cabinet most often, and they are very familiar with the shape and size of the express parcel in the delivery area. Therefore, the courier is mainly used as the object of investigation.

When designing the questionnaire, the size of the express parcel must be considered to provide a reference for the survey respondents. The application scope of national standard express carton boxes is relatively wide, the size and specifications are relatively complete, and the courier is also very familiar with the situation of express carton boxes in daily delivery. Therefore, the national postal standard courier box is used as a reference to design the questionnaire for the questionnaire survey. In the questionnaire, express parcel is divided into four categories: file type, small-sized, middle-sized and large-sized. The survey area is divided into three types: communities, office buildings, and universities. From each region, 150 copies of data were collected, and the data style was proportional data of different types of express parcels, such as file type: small-sized: middle-sized: large-sized = 1: 3: 4: 2. A total of 450 data were received from the three regions. Due to different regions, there may be different differences in consumption habits and consumption characteristics, so the situation of various shipments in different regions is compared to analyze whether there are differences.

2) *Data collection and statistics:* Descriptive statistical analysis is a basic analytical process. Use mathematical calculations to reflect the overall characteristics of statistics, including indicators such as distribution status, numerical characteristics, and special values. By describing the statistical analysis, it can understand the degree of concentration and dispersion of the data, and explain whether the respondents' perception of the survey questions is consistent or there is a large deviation, which is the basis of cluster analysis.

3) *Comparison of file type packages:* Among the communities, 68.7% of the communities' file type packages accounted for about 10% of the parcels, and 30% of communities' file type packages accounted for about 20% of the parcels. Among office buildings, file type packages in 65.3% of office buildings accounted for more than 30% of the parcels, and file type packages in 34.0% of office buildings accounted for 32.3% of the parcels. Among universities, the file type parcels in 62.7% of universities account for about 10% of parcels, and the file type parcels in 32.3% of universities make up about 20% of parcels. From the

distribution of data, the file type packages in more than 95% of communities and universities account for about 10% -20% of the packages. The communities are mainly a place of living. The main consumer group of universities is students, and most of the consumption characteristics of students are daily necessities and school supplies, so file types packages in these two types of regions account for a small proportion. Among all the office buildings surveyed, file types packages accounted for 30% -40% of all courier services, which is quite different from communities and universities. The office building is an office space, with various documents, contracts, and bills, so the file types packages account for a large proportion.

B. Comparison of Small-sized Packages

Among the communities surveyed, small-sized parcels accounted for about 20% in 66% of the communities, and small-sized parcels accounted for about 30% in 28.7%. Among the office buildings, about 20% of the small-sized parcels are in 28.7% of the office buildings, and about 30% of the small-sized parcels are in 70.7%. In universities, small-sized parcels in 90% of universities are basically distributed between 40% and 50%. The distribution of personnel in the community is relatively complicated, and their income levels are different. Therefore, small-sized parcels are mainly small-sized living and household items. The office building is an office space, and in addition to documents and contracts in the area, there are also small office supplies for the company. In universities, student consumption is mainly for living and school supplies. The proportion of small-sized express delivery is slightly higher than that of community and office buildings.

C. Comparison of Middle-sized Packages

In the community, middle-sized packages accounted for about 40% in 65.3% of the communities. Among office buildings, middle-sized packages accounted for about 20% of 56.7% office buildings, and middle-sized packages accounted for about 30% of 42.7% office buildings. Among universities, middle-sized packages in 68% of universities account for about 40%. The community is a place of life, and the purchases of college students are also daily necessities, so the distribution of middle-sized packages in these two areas is basically similar. While office buildings are office spaces, middle-sized packages are slightly less than that in the other two types of areas.

D. Comparison of Large-sized Packages

In the community, large-sized packages account for about 30% of the 60.7% community; large-sized packages are generally distributed between 10% -20% in office buildings and universities. In the community, customers may buy some small appliances or a lot of daily necessities, so the proportion of large-sized packages is higher than that of the other two regions. The office building may have some slightly larger office supplies, and the items that students buy are generally basic supplies for study or living, so there are not many large-sized packages.

E. K-means Cluster Analysis

The K-means method is to first select K sample data from all data randomly as the initial cluster center; secondly, according to the distance measurement standard, the distance between all the data and the initial clustering center is calculated and classified into the nearest clustering center to form K-type data; then, the average of each type of data is calculated to be the new cluster center point. Then, it is needed to continue to calculate the distance from each sample to the new cluster center point, and reclassify all the data. This calculation is repeated until the cluster termination condition is satisfied. There are two kinds of termination conditions to choose from: one is to meet the set number of repeated iterative times, e.g. the system default is 10 times, and the cluster is terminated after reaching the number of times. The second is the distance between the new cluster center and the previous cluster center. The system defaults it to 0.02. When the distance between the new class center and the previous class center is less than the set value, the cluster terminates.

The K-means algorithm is used to statistically analyze the data of the express size ratio in all regions, and the number of clusters $K = 3$ is selected for calculation. The iterative times number is the system default 10 times. The clustering process is shown in the figure. In K-means clustering, the starting cluster center is randomly selected, the final cluster center is formed through iterative operations, and all data are classified into the nearest cluster. By clustering the data of the express size ratio in different regions, it can get the starting cluster center table, the final cluster center table, the iterative process, the distance table to the final cluster center, the member table of each cluster, the number of observations in the new cluster center, and the area number after the aggregation. Based on data summary statistics, "Table I" shows the starting cluster center and the final cluster center when K-means clustering, and "Table II" shows the number of observations in each cluster center when the number of clusters is $K = 3$.

TABLE I. COMPARISON OF CLUSTERING CENTERS

	Starting Cluster Center			Final Cluster Center		
	Office buildings	Communities	Universities	Office buildings	Communities	Universities
File type	5	2	1	4	1	1
Small-sized	2	3	4	2	2	4
Middle-sized	2	5	3	2	4	3
Large-sized	1	2	2	2	3	2

TABLE II. OBSERVATION DATA FOR EACH CLUSTER CATEGORY AT $K = 3$

Area	Category		
	Category I	Category II	Category III
Communities	2	130	18
Office buildings	148	0	2
Universities	0	36	114

According to "Table I" and "Table II", during K-means clustering, the initial clustering center randomly selects data, and after certain iterative times, the final clustering center is formed, and 450 data are divided into three categories. Category I consists of 2 communities and 148 office buildings. Basically, all office buildings are grouped into one category. The final cluster center ratio is: file type: small-sized: middle-sized: large-sized = 3: 3: 2: 2. Among them, file type and small-sized packages are more, while middle-sized and large-sized packages are relatively less. Category II consists of 130 communities and 36 universities, 87.1% of communities and a small number of universities are clustered into one category. The final cluster center ratio is: file type: small-sized: middle-sized: large-sized = 1: 2: 4: 3, where file type accounts for the smallest proportion, middle-sized accounts for the largest proportion, small-sized and large -sized is also more than file type. The community is a place of living. The main types of

courier are daily necessities, clothing, shoes and hats, small furniture appliances, and mother and baby products, so middle-sized and large-sized account for more. Category III consists of 114 universities, 18 communities, and 1 office building, 75.6% of universities and a small number of communities are clustered into one category. The final cluster center ratio is: file type: small-sized: middle-sized: large-sized = 1: 4: 4: 1. The consumer group of universities is mainly students, and the types of students' consumption are mainly small and middle-sized items such as food, clothing, shoes and hats, daily necessities, books, etc. Therefore, small-sized and middle-sized packages account for a large proportion, and file type and large-sized packages also have a certain proportion.

Because the methods of calculating distances between different algorithms are slightly different, the results obtained by system clustering and K-means clustering have a certain degree of overlap when clustering between communities and universities. But most of them have some differences, and are generally similar. Basically, all office buildings belong to one category, 60% to 80% of communities belong to one category, and 70% to 90% of universities belong to one category. In addition, there are obvious differences in the proportion of express parcels between different regions, which should be based on the final cluster center, as shown in "Table III":

TABLE III. PROPORTION OF PARCEL SIZE BETWEEN DIFFERENT REGIONS

Area type	file type	Small-sized	Middle-sized	Large-sized
Communities	1	2	4	3
Office buildings	4	2	2	2
Universities	1	4	3	2

IV. CONCLUSION

This study obtains the proportion of courier parcels by investigating the proportion of courier parcels in different types of regions, and by surveying the delivery situation of couriers in three different types of communities, universities, and office buildings. The K-means clustering method of SPSS software was used to perform cluster analysis on the collected data, and finally the size ratio of the express cabinets in different types of locations was optimized. By optimizing the size and configuration ratio of the cabinet, the overall turnover rate and customer satisfaction of the express cabinet are improved.

REFERENCES

- [1] Liu Lihua. Application and Development Direction of Self-service of Intelligent Express Cabinet [J]. *Logistics Engineering and Management*, 2015,08. (in Chinese)
- [2] Du Rongxue, Ruan Guoxiang. Feasibility Analysis of The Application of Intelligent Express Cabinets in Urban Communities [J]. *Logistics Science and Technology*, 2015,01. (in Chinese)
- [3] Lei Xiaoqing, Wang Qian. Development Experience of US-Japan Online Shopping Self-Pickup Distribution Model and Its Enlightenment to China [J]. *Logistics Science and Technology*, 2014, 11. (in Chinese)
- [4] Li Yue. Research on Optimal Combination of Strategies for Improving Turnover Efficiency of Intelligent Express Cabinets Based on Orthogonal Design [J]. *Modern Business*, 2017, 05. (in Chinese)