Clustering-Based Analysis of E-commerce Customers' Consumption Behavior in the Post-epidemic Period

Zhengyan Cui(✉)

School of Management, Shanghai University, Shanghai, China
czyan0315@163.com

Abstract. The COVID-19 epidemic in 2020 brought huge changes to the world, causing many people to engage in online shopping and some consumers' purchasing behavior to change before and after the epidemic. In order to better study the consumption behavior of e-commerce customers in the post-epidemic era, this study takes Chinese online shoppers as the research target and uses questionnaires to collect their consumption behavior preferences and to quantify them. This article conducts K-means cluster analysis on consumer behavior data in the post-epidemic era to study the characteristics of different types of consumer behavior, classifying consumers into general consumers, purchase consumers, motivated consumers, product consumers and evaluation consumers. I also make targeted recommendations to e-commerce platforms to help them and hope these will help e-commerce platforms to better cope with the impact of the COVID-19 and enable them to retain customers and achieve accurate marketing in the post-epidemic era, which will ultimately benefit both buyers and sellers.

Keywords: Consumer behavior · E-commerce customers · Cluster analysis

1 Introduction

With the global outbreak of the COVID-19 epidemic in 2020, it has exposed many companies to turmoil and danger. Many offline shops have closed down due to lack of customers, accompanied by a boom in online shopping [1]. With the non-contact demand of the COVID-19 epidemic, as well as the mass availability of the internet and the simplicity of consumer perception, it is clear online shopping has become the current mainstream form of consumption [2]. Consumers' reliance on e-commerce shopping has gradually deepened, further contributing to the booming online shopping market. Until 2021, China's online retail transactions have exceeded 13 trillion yuan.

V. Gaikar et al. (Eds.): ECIT 2023, AHE 18, pp. 30–35, 2023.
https://doi.org/10.2991/978-94-6463-210-1_5

2 Research Methodology and Design

2.1 Research Methodology

1) **Cluster analysis method.**
Cluster analysis, also called classification analysis or numerical classification, is a multivariate analysis method in statistics to study the problem of "clustering of things by classes" [3], which can automatically classify a group of observations (or variables) according to their many characteristics and according to their closeness in nature without a priori knowledge, producing multiple classification results. The similarity of individual characteristics within a class and the variability of individual characteristics across classes are large [4].
2) **Measure of "closeness" in cluster analysis.**
"The smaller the distance between points, the "closer" they are, and the more likely they are to cluster together. The greater the distance between points, the further they are from each other and the less likely they are to cluster together [5].
Let the eigenspace χ be the n-dimensional real vector space Rn, $x_i, x_j \in \chi$, where:

$$x_i = \left(x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(n)}\right)^T \tag{1}$$

$$x_j = \left(x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(n)}\right)^T \tag{2}$$

Then the distance between the points is (where p $\geq$ 1).

$$L_p(x_i, x_j) = \left(\sum\nolimits_{l=1}^{n} \left|x_i^{(l)} - x_j^{(l)}\right|^p\right)^{\frac{1}{p}} \tag{3}$$

When p = 1, it is the Manhattan distance:

$$L_1(x_i, x_j) = \left(\sum\nolimits_{l=1}^{n} \left|x_i^{(l)} - x_j^{(l)}\right|^1\right)^{\frac{1}{1}} = \sum\nolimits_{l=1}^{n} \left|x_i^{(l)} - x_j^{(l)}\right| \tag{4}$$

When p = 2, it is the Euclidean distance:

$$L_2(x_i, x_j) = \left(\sum\nolimits_{l=1}^{n} \left|x_i^{(l)} - x_j^{(l)}\right|^2\right)^{\frac{1}{2}} \tag{5}$$

When p = ∞, it indicates the maximum value of each coordinate distance:

$$L_\infty(x_i, x_j) = max\left|x_i^{(l)} - x_j^{(l)}\right| \tag{6}$$

In this article, the Euclidean distance is used.

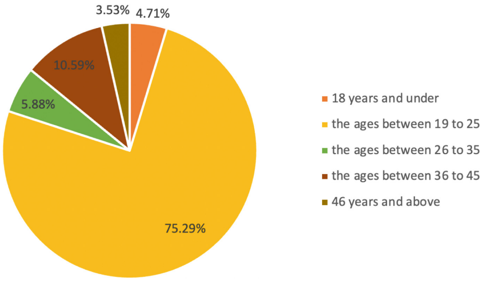

Fig. 1. Age distribution pie chart.

3 Results

3.1 Descriptive Analysis of Data

1) **Age distribution of survey respondent.**

From the age distribution of survey respondents (Fig. 1), the largest number of respondents are between 19–25 years old, accounting for 75.29%; followed by 35–45 years old, accounting for 10.59%; the remaining age groups are less distributed, with 5.88% of people between 26–35 years old; only 4 people are over 46 years old, and 5 people are 18 years old or younger; from this we can see that The respondents are mainly concentrated in the group between 19–45 years old, which is more in line with the age of the group I envisioned for online shopping, and online shopping is more common and accepted in this group.

2) **Income of survey respondent.**

From the income distribution of the surveyors (Fig. 2 and Fig. 3), the majority of surveyors have a monthly income level below 2,500 yuan, accounting for 65.88%; followed by 2,500–5,000 yuan accounting for 18.82%, monthly income of 5001–10,000 yuan accounting for 12.94%; the number of 10001–15,000 yuan and 15,000 yuan or more, respectively, 2 people, 1 person The number of people with monthly income of 5001–10000 yuan reached 12.94%; the number of people with monthly income of 10001–15000 yuan and 15000 yuan or more were 2 and 1 respectively. It can be seen most of the respondents are low-income and middle-income people, and there are fewer high- income people.

3.2 K-means Clustering Analysis

1) **Determination of k-value.**

The determination of K-value is important in the K-means algorithm [6]. In this article, the elbow method is used to determine the K-value. The core idea of the elbow method is that as the K value increases, the samples will be divided more precisely and the degree of aggregation of each cluster will increase, then the SSE (sum of squared errors) will become smaller and smaller [7]. When the value of K does not reach the optimal clustering, the increase of K will greatly increase the degree of aggregation of each cluster, so the SSE will decrease greatly [8]. And when K reaches the optimal number of clusters, the degree of aggregation obtained by increasing K will decrease,

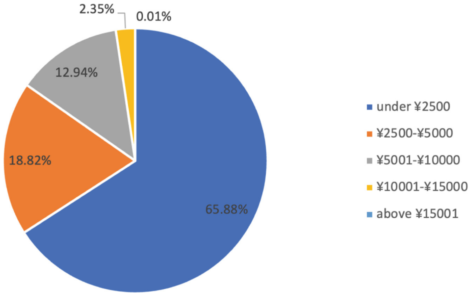

Fig. 2. Income distribution pie chart.

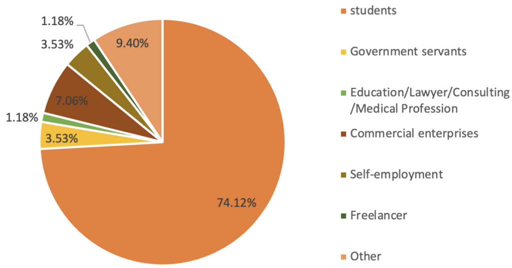

Fig. 3. Occupational distribution pie chart.

so the decrease in SSE will decrease sharply and then flatten out as the value of K continues to increase [9]. Graphically, SSE and K are like the shape of an elbow, and the value of K corresponding to this elbow is the optimal number of clusters. In this article, the graph of SSE versus K value was obtained using Python (as in Fig. 4). Therefore, the K value of this experiment is taken as 5.

2) **K-means clustering.**

K-means clustering was performed using SPSS, and the following results were obtained by inputting 5 in the number of clusters [10].As show in Table 1.

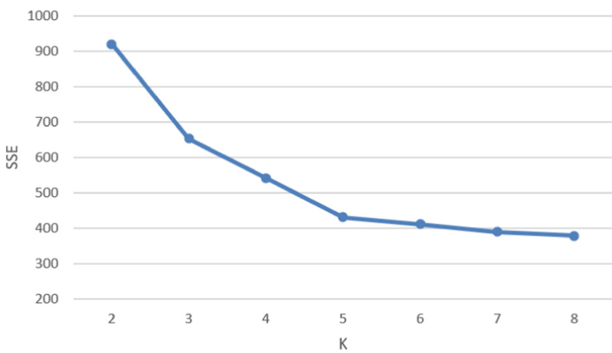

Fig. 4. SSE versus K value.

Table 1. Initial clustering centers.

	Clustering					
	1	1	2	3	4	5
Consumer products	5	5	6	5	2	6
Consumption frequency	1	5	4	2	2	3
Consumer motivation	3	3	9	12	4	7
Purchase behavior	8	8	17	9	13	12
Payment behavior	1	1	2	2	3	2
Post-purchase behavior	2	2	3	5	2	5

After obtaining the clustering centers, the Euclidean distance from each point to the initial clustering center was calculated and iterated continuously until the requirements were met and then the iteration was stopped.

4 Conclusion

In the past, most e-commerce platforms paid more attention to the technical optimization and improvement of the platform, while ignoring the grasp of consumer behavior. In today's post-epidemic era, consumer behavior is changing subconsciously under the influence of the environment. Under the impetus of big data, user data has become the core resource of e-commerce platforms, including consumer behavior data, which can help e-commerce platforms enhance their core competitiveness and consolidate their competitive position if they are carefully analyzed and studied for their intrinsic regularity. Therefore, it is of strong practical significance for e-commerce platforms to study consumer behavior data. Therefore, this thesis uses questionnaires to investigate the consumption behaviors of different consumers when shopping on e-commerce platforms after the epidemic, study the characteristics of their consumption behaviors, and perform k-means cluster analysis to classify consumers' consumption behaviors after the epidemic into five categories: general, purchase, motivation, product, and evaluation consumers.

References

1. Xin, X.L.; Guo, C.L. A study on consumers' shopping intention on cross- border e-commerce platform under the epidemic situation. China Forestry Economy, 2022, 5, 75-81.
2. Chandron, R.; James, F.; Roger, D.; et al. Consumer Behavior. Hinsdale. Journal of Advertising, 1979, 8, 52~53.
3. Ajzen, I.; Fishbein, M. Attitude-behavior relations: A theoretical analysis and review of empirical research. Psychological Bulletin, 1977, 84, 888~918.
4. Suleymanov, F. Consumer behaviour and agricultural consumer market. Scientific Papers-Series Management Economic Engineering in Agriculture and Rural Development, 2021, 21, 753-758.

5. Luck, M.; Benkenstein, M. Consumers between supermarket shelves: The influence of interpersonal distance on consumer behavior. Journal of Retailing and Consumer Services, 2015, 26, 104-114.
6. Reed, A.; Forehand M. R.; Puntoni, s.; et al. Identity-based consumer behavior. International Journal of Research in Marketing, 2012, 29, 310–321.
7. Beick, D. J.; Wight, K. G.; fitzsimons, G. J. Secret consumer behavior in close relationships. JOURNAL OF CONSUMER PSYCHOLOGY, 2022.
8. Dennis, C.; Merrilees, B.; Jayawardhena, C.; et al. E-consumer behaviour. European Journal of Marketing, 2009, 43, 1121-1139.
9. Liu, X.; Kwok, K. W. An empirical study of product differences in consumers' E-commerce adoption behavior. Electronic Commerce Research and Applications, 2003, 2, 229-239.
10. Huang, W. L.; Hu, P.; Tsai, S.; et al. The business analysis on the home-bias of E-commerce consumer behavior. Electronic Commerce Research, 2020, 21, 855-879.https://doi.org/10.1007/s10660-020-09431-2