J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 6711--6728

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on Cross-border E-commerce Consumer Behavior Based on Spatio-temporal Data Mining Calculation in the Perspective of Digital Economy Globalization

Yingchao Lu[1,✉], Sijia Lv[2]

[1] *School of Management, Seoul School of Integrated Sciences & Technologies (aSSIST University), Seodaemun, Seoul, 03600, Korea*

[2] *Harbin Medical University Cancer Hospital, Harbin Medical University, Harbin, Heilongjiang, 150081, China*

ABSTRACT
Thanks to the wave of digital economic globalization, the business development of cross-border e-commerce platforms is in full swing. This paper aims to promote the development of e-commerce personalization and launch the research of consumer behavior characteristics. This paper utilizes the concept of entropy in information theory to modify the weights of user feature vectors, so as to make up for the inadequacy of the K-Means algorithm in expressing ambiguous clustering information. Combined with the data samples, the consumer behavior prediction model is established. For the dynamic clustering of customer groups, construct the customer segmentation model based on the improved K-Means algorithm. Combined with the time series prediction model, complete the formation of the spatio-temporal data mining model of consumer behavior. The model is used to mine the consumer behavior dataset of a cross-border e-commerce platform, and the clustering analysis yields four precise consumer group portraits. In this paper, by mining and analyzing the characteristics of consumer spatio-temporal data, the cross-border e-commerce platform is provided with more accurate user insights and marketing optimization solutions.

Keywords: K-Means algorithm; cross-border e-commerce; spatio-temporal data; consumer group portrait

1. Introduction

With the rapid development of information technology and the advancement of digital transformation, digitalization has brought far-reaching impacts and great changes to the business environment and consumer behavior. In the digital era, consumer behaviors and expectations are

✉ Corresponding author.
E-mail address: yingchaolu1996@163.com (Y. Lu).
Received 05 February 2024; Revised 12 April 2024; Accepted 05 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-367

rapidly evolving, and e-commerce is changing the face of business at an unprecedented rate [1]. People not only use digital technologies in their daily lives, but also widely apply them in shopping, socializing, and entertainment, and this digital lifestyle is profoundly shaping consumer mindsets, behaviors, and expectations [2-4]. Traditional marketing and business models are no longer adapted to this digital world, and there is an urgent need to explore the evolution of consumer behavior and e-commerce trends in depth so that we can better understand and respond to this digital era full of opportunities and challenges, and provide strategic guidance and management ideas for the sustainable development of enterprises [5-7].

Among them, cross-border e-commerce is a product of the combination of Internet technology and globalization process [8]. It reduces trade barriers, improves the convenience of international trade, and provides global consumers with more choices of goods and services [9-10]. Based on this, consumers' shopping behavior has changed significantly. For example, consumers are more and more inclined to online shopping, their reliance on e-commerce platforms is increasing, and consumers' shopping decision-making process has become more complex [11-12]. These changes undoubtedly put forward higher requirements for cross-border e-commerce platforms. Therefore, an in-depth study of the transformation of consumer shopping behavior under the development trend of cross-border e-commerce is of great significance for the healthy development of cross-border e-commerce [13-14].

By mining and analyzing the platform user behavior data, it predicts the platform user consumption behavior. Combined with the consumption behavior dynamic segmentation of customer groups, further clustering analysis to get the consumer portrait, so as to provide optimization and improvement reference for the operation of cross-border e-commerce platform. Under this research idea, this paper firstly discusses the construction and functional implementation of the consumer behavior prediction model, focusing on the improvement of the K-Means clustering algorithm by the entropy concept. Subsequently, the customer dynamic clustering model is constructed to extract user groups with different attribute behaviors. Combined with the time series prediction model, the spatio-temporal data mining model of consumer behavior is created. The model is used to mine the consumer behavior dataset of a cross-border e-commerce platform, analyze the characteristics of different consumer behaviors and perform clustering analysis. Based on the clustering results, consumer profiles are drawn, and corresponding marketing strategy suggestions are given based on the characteristics of different profiles.

2. Consumer behavior prediction models

In this paper, the consumer behavior prediction model is divided into two modules: data sample and mining algorithm. This chapter firstly describes the source of data samples and basic classification, and then chooses K-Means clustering algorithm to mine consumer behavior features. Finally, the K-Means clustering algorithm is optimized for the lack of feature description, so that the performance of the consumer behavior prediction model is further improved.

2.1. Data sources

Mobile and PC can access information at any time, and the scenario information is richer, such as user location information and user access time. At present, cross-border e-commerce relies on the power of the network, thus crossing the barriers of time and space in the traditional cross-border sales methods, and consumers can choose merchants without restriction. Cross-border e-commerce platforms can improve consumer satisfaction by collecting information on consumer behavior, tapping into consumers' different platform demands, and accurately recommending goods of their interest to users in real time. Obtaining user interest preferences, which in turn improves user activity and purchase conversion rate. At the same time, the cross-border e-commerce platform uses clustering and other means to analyze the consumer data on the platform, categorize consumers, find out which is the potential for consumers who have not purchased, and use targeted push offers and so on to revitalize this part of the potential stock, convert potential consumption into habitual consumption, and achieve the overall penetration of cross-border mall goods.

2.2. Feature Selection and Engineering

In consumer behavior prediction models, feature selection is the core element to guarantee the accuracy of the model. Specifically, features can be roughly divided into three categories: user features, product features and interaction features. User features include user's age, gender, geographic location, and historical purchase behavior. Product features include key information such as product price, category, brand, and rating. Interaction features focus on the interaction data between users and products, such as the number of clicks, purchase frequency and browsing time.

In order to further improve the model performance, feature engineering is especially critical. In feature engineering, multiple methods are usually used to optimize the original features to improve the prediction effect of the model. For example, through feature transformation, normalization or standardization can be applied to improve the distribution or scaling of features. Feature combination, on the other hand, creates new features by combining two or more features to capture more information.

2.2.1. Feature weights. Features are the information that describes a sample, but there will be redundancy of information and irrelevance of information in the feature dimension, these factors will bring some error in the model, so studying the features and selecting the subset of features that can be appropriate from the features has a very important role in the data mining project.

2.2.2. Feature engineering. Throughout the existing mathematical modeling competitions, the participating champion teams often did not use very profound and complex mathematical models, and most of them achieved good performance by doing sufficient research in feature selection as well as processing, and then using some simpler models, such as algorithms of logistic regression, linear regression, and so on. In addition, the selection of high-quality features can enhance the generalization ability of the model and the speed of operations, making the model more interpretable and easier to maintain.

Regrettably, in most data mining books and papers, optimization and improvement are only carried out at the algorithmic level, and feature engineering is not mentioned, which leads to novices doing data mining to pursue the complexity of the algorithm and ignore the research on features.

A complete feature engineering needs to do work is shown in Figure 1.

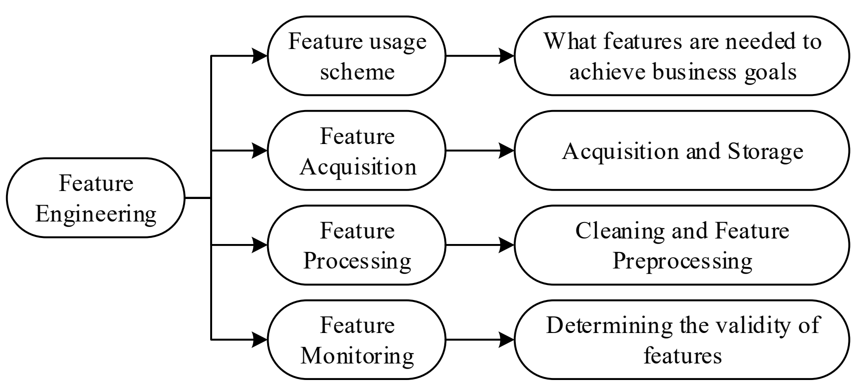


Figure 1. Feature engineering

2.2.3. Information entropy. Any objective things need information to describe, information is an effective way to express the characteristics of things, it ensures that anything or the objective world has a certain unique internal existence of the structure or mechanism. Information in the objective world is diverse, different combinations of information will express different things, information is the door to intuitive understanding of the world, effective information can enhance people's knowledge of things, which can be achieved to reduce the uncertainty of things. Information theory is such a science, information theory is mainly based on probability theory and mathematical statistics, from the quantitative point of view, qualitative or quantitative description of things, it is from an objective point of view to study the need for information measurement, acquisition, etc., in order to solve the communication field of data transmission, data compression and other related issues in the field of discipline. Initially, this discipline was proposed by scholars to solve the problem of effective information transfer in data communication. At that time, it was mainly used to solve the common problems of measurement, exchange and transmission in communication engineering. However, with the deepening of theoretical research in basic disciplines and the increase of relevant cases, information theory and information science have been improved, and scholars have deepened their understanding of information theory, and now the scope of research has been extended to all fields related to information. For example, linguistic disciplines, psychological research disciplines, etc., among which, data mining is also a field it involves.

The concept of entropy is originally derived from the discipline of statistical thermodynamics, which indicates the degree of uncertainty of a system, where the higher the uncertainty of the system, i.e. the more chaotic the system is, the higher the entropy value of the system is. On the contrary, if the state of the system is stable, then the entropy value corresponding to this system is lower. In the study of information theory, entropy is usually and often referred to as information entropy, which has the main advantage of expressing the degree of uncertainty of a random variable quantitatively in numerical form, with the ultimate goal of visualizing the amount of information content.

Assuming that the probability density function of X is $P(x)$, the information entropy of the distribution is Eq. (1):

$$H(x) = -\int P(x)\log P(x)dx \tag{1}$$

From equation (1), we can know that the value of $H(x)$ is only related to the distribution of X. When all the values of X have the same probability, then the entropy value of X is the largest, which means that the uncertainty of X is the largest. Because at this time, there is no way to determine the value that X takes. On the contrary, if the variable X takes only one value x, then the probability distribution of that value $P(x)$ is 1. At this time, X is completely certain, and its entropy value reaches the minimum.

Assume that x is a discrete random variable taking finite values, and its probability distribution is equation (2):

$$P(X = x_i) = p_i \tag{2}$$

Then the entropy of the random variable X is defined as equation (3):

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \tag{3}$$

From the above definition, it can be seen that the entropy depends only on the distribution of X and has no relationship with the value of X. Therefore, the entropy of X can also be written as $H(p)$ as in equation (4):

$$H(p) = -\sum_{i=1}^{n} p_i \log p_i \tag{4}$$

When the random variable takes only two values, e.g., 1, 0, i.e., the distribution of X is equation (5):

$$P(X=1)=p, P(X=0)=1-p \tag{5}$$

The entropy is then equation (6):

$$H(p)=-p\log p-(1-p)\log(1-p) \tag{6}$$

3. Spatio-temporal data mining model of consumer behavior

From the above, it can be seen that the consumer behavior prediction model can objectively analyze the platform user behavior, so as to realize the prediction of consumer behavior. And the proposed improved K-means algorithm can fully consider the differences between the sample features, under this premise, this chapter establishes a customer segmentation model based on the improved K-Means algorithm to realize the fine segmentation of different consumers. Combined with the time series prediction model, it realizes the spatio-temporal data mining of consumer behavior.

3.1. Dynamic customer segmentation model based on improved K-Means algorithm

Aiming at the shortcomings of existing customer segmentation methods, this paper proposes a dynamic RFM customer segmentation model based on improved K-Means clustering to obtain the user group to which each user belongs every week, and accordingly derive the transfer probability matrix between user groups to accurately portray the change of user loyalty and the transfer characteristics between user groups. The RFM model is used as an important tool and means for evaluating the value of users, and its R denotes the time of the user's latest consumption, F denotes the frequency of the user's consumption in the recent period of time, and M denotes the amount of the user's consumption in the recent period of time.

The model is divided into a total of the following four steps:

1) The RFM model based on the entropy weight method and the customer segmentation model based on K-Means are combined to determine the user group division criteria. Since each user has a different survival period on the platform (survival period is the number of weeks from the user's first purchase to the 15th week of 2020), it is necessary to utilize the survival period weighting in the calculation of the indicators of the RFM, which is calculated as follows:

First, the total F (frequency of purchases) and M (amount of purchases) for each user during the observation period were converted to MAF (average weekly frequency of purchases) and MAM (average weekly amount of purchases) and R (number of weeks since last purchase to the 15th week of 2020) by utilizing survival weighting to obtain each user's MAF, MAM, and R metrics.

Second, due to the different units, the above MAF, MAM, and R indicators are normalized to eliminate the effect of magnitude. Here, min-max normalization is used to normalize the data into the $[0,1]$ interval to obtain the normalized indicators, i.e., N-MAF, N-MAM, and N-R indicators.

Then, based on the entropy weighting method to determine the weights for the above normalized processed indicators, the weighted indicators are obtained, i.e., W-MAF, W-MAM, W-R indicators.

Finally, K-Means clustering algorithm is used to cluster the weighted indicators, and CHI, contour coefficient and inertia indicators are used as the evaluation standard of good or bad clustering to get the number of optimal user group divisions and the corresponding clustering centers.

2) Dynamic calculation of consumer MAF, MAM and R metrics. Calculate the weekly MAF, MAM, and R metrics for each user by using the sliding window approach, and the logic of calculating the metrics for user i in week t is as follows:

Calculate the MAF (Average Weekly Purchase Frequency), MAM (Average Weekly Purchase Amount), R (Number of Weeks Since Last Purchase in Week t) for user i in the $[t-3,t]$ weekly window and min-max normalization, accordingly, we can get the N-MAF, N-MAM, N-R metrics of each user in each week. , N-R metrics, and then using the metric weights set in step (1), the weighted metrics, i.e., W-MAF, W-MAM, W-R, are obtained.

3) Dynamic consumer clustering. Based on the weekly W-MAF, W-MAM, and W-R metrics per user obtained in step 2) and the clustering centers determined in step 1), the Euclidean distance is used as a metric to dynamically tag users on a weekly basis, determine the user group to which the user belongs on a weekly basis, and obtain the user state matrix.

4) Consumer group migration probability matrix. Based on the user state matrix obtained in step 3), calculate the one-step transfer probability matrix between user groups. The specific calculation logic is as follows:

First calculate the probability of each state a_i at the moment of t as in equation (7):

$$P\{X_t = a_i\} \tag{7}$$

Calculate the probability that the state at the moment of t is a_i and the state at the moment of $t+1$ is a_j as in equation (8):

$$P\{X_t = a_i, X_{t+1} = a_j\} \tag{8}$$

Calculate the transfer probability as in equation (9):

$$P_{ij} = P_{ij}(l) = P\{X_{t+1} = a_j \mid X_t = a_i\} = P\{X_t = a_i, X_{t+1} = a_j\} / P\{X_t = a_i\} \tag{9}$$

where l is the state space.

The design steps of the dynamic RFM customer segmentation model based on improved K-Means clustering involved in this paper are shown in Fig. 2.

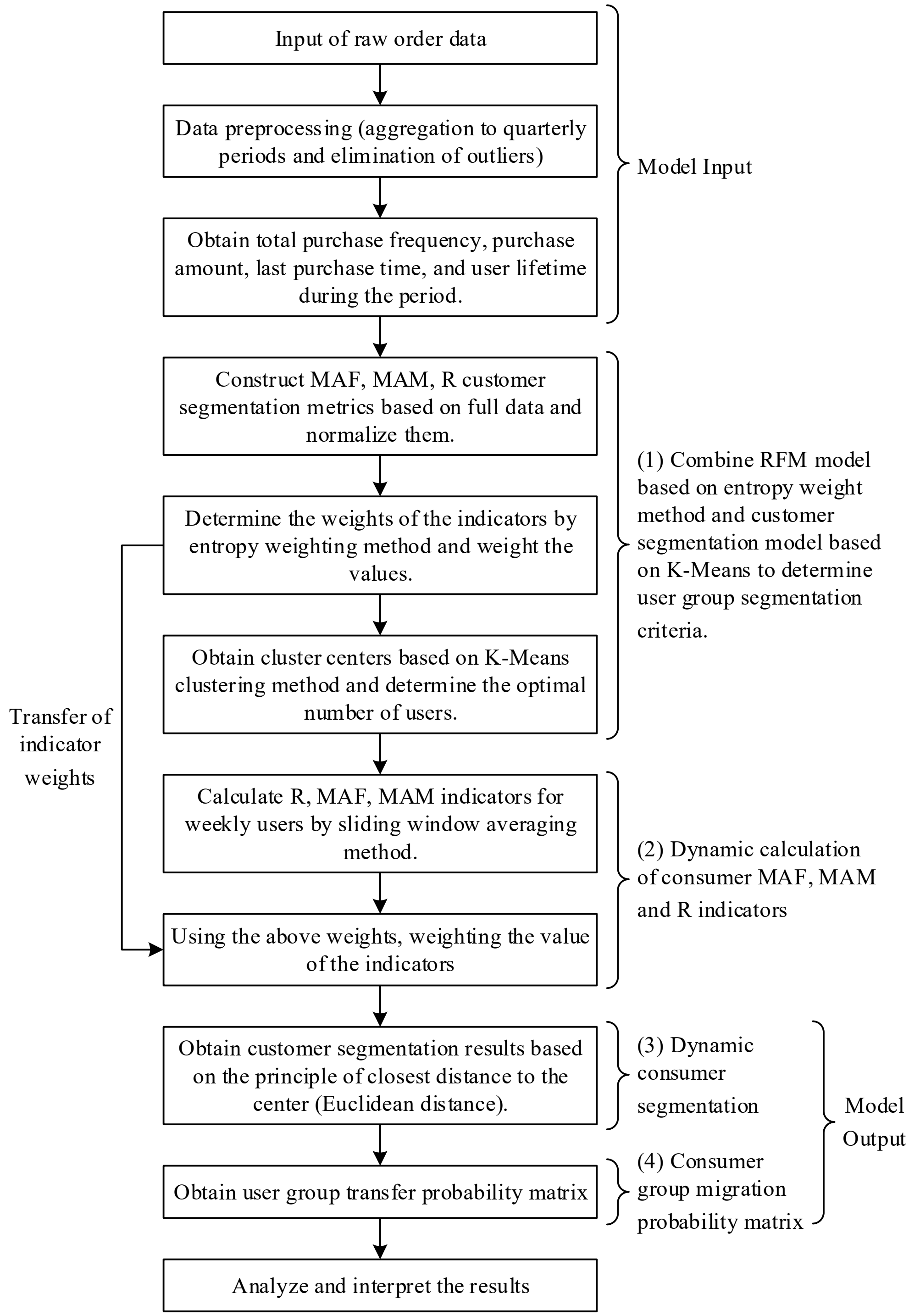


Figure 2. Dynamic RFM customer segmentation model design steps

3.2. Time series forecasting models

Using a time series forecasting model, an interpretable forecast can be given from the perspective of historical data by fitting different categories of products, and factors such as sales trends, sales cycles, seasonal variations, and regular promotional activities are already embedded in such a forecasting model. However, the prediction results based on the time series forecasting model have certain limitations, first of all, such a prediction value can not be brought into the historical differences and other factors, for the annual cycle of sales forecasts, such as this year's lower than last year's temperatures, promotional efforts such as such factors can not be a better response in the time series model. Currently the mainstream time series forecasting methods: moving average (MA), exponential smoothing (ES), integrated moving average autoregressive method (ARIMA).

1) Moving Average. There is a time series X_n composed of n observations where X_t is the data of the first t period $(t=1,2,\cdots,n)$, and the data of the consecutive N observation periods ($N<n$, called the time interval) are selected to calculate the moving arithmetic mean M_t as the predicted value as in Eq.(10) :

$$M_1(t)=\frac{X_t+X_{t-1}+\cdots+X_{t-N+1}}{N} \tag{10}$$

Because $M_1(t)$ has a lag bias, $M_1(t)$ is essentially an arithmetic average of the past few data, and there is $M_1(t)<Y_t$ when the time series is trending up, and $M_1(t)>Y_t$ when the time series is trending down. Therefore, when using one-shot moving average method for forecasting, especially for trending time series, the predicted values need to be shifted forward by $\rho=(N-1)/2$ positions in order to track the actual trend.

2) Exponential Smoothing. Exponential smoothing is a special kind of weighted moving average method, which has the advantages of continuous application, no need to retain historical data, easy calculation, easy to update the forecast model, so it is a commonly used market forecasting method. The practical application of exponential smoothing can be categorized into primary exponential smoothing and multiple exponential smoothing.

With a time series X_n composed of n observations where X_t is the data of the first t period $(t=1,2,\cdots,n)$, the formula for calculating the value of one-time exponential smoothing is equation (11):

$$S_1(t)=\alpha X_t+(1-\alpha)S_1(t-1) \tag{11}$$

where:

$S_1(t)$ - the value of the primary exponential smoothing in period t.

α - the smoothing coefficient $(1-\alpha)$ called the damping coefficient.

$S_1(t-1)$ - the value of primary exponential smoothing in the $t-1$ period.

According to the derivation of the formula it can be shown that given the smoothing coefficient α and the initial value $S_1(0)$, it is possible to calculate the value of primary exponential smoothing. Expansion leads to equation (12):

$$S_1(t)=\alpha X_t+\alpha(1-\alpha)X_{t-1}+\cdots+\alpha(1-\alpha)^{t-1}X_1+(1-\alpha)^t S_t(0) \tag{12}$$

It can be found that the sum of the coefficients on the right-hand side of equation (12) is 1. Therefore, the one-time exponential smoothing value is essentially a weighted average of X_n, $S_t(0)$, and the recent data are given more weight, reflecting the idea of emphasizing the recent data. Since the exponential smoothing value is a weighted average of the historical data, it has a lagging deviation relative to the original series. Therefore, using the one-shift moving average method for forecasting, the predicted value needs to be shifted forward by $\rho=(1-\alpha)/\alpha$ positions in order to track the actual trend.

3) ARIMA. ARIMA forecasting method, also known as Box-Jenkins model, is a short-term time series forecasting model with high accuracy. The basic idea of ARIMA method is: the formation of the sequence of forecasting objects over time, as a series of time-varying and interrelated digital sequences, which can be described by appropriate mathematical models. ARIMA model does not construct individual equations or co-equation models, but rather analyzes the probabilistic or stochastic nature of economic time series and lets the data speak for itself. There are three basic

forms of the model: the autoregressive model (AR), the moving average model (MA) and the mixed model (ARIMA).

A time series $\{X_t\}$ is said to obey an integrated moving average autoregressive model of order (p,q), denoted as $ARIMA(p,q)$, if the time series $\{X_t\}$ satisfies Eq. If $q=0$, the model is $AR(p)$. If $p=0$, the model is MA(q). The series to which the ARIMA model applies should be smooth series, with equation (13):

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon \tag{13}$$

Where: $\{\varepsilon_t\}$ is a sequence of independent identically distributed random variables and satisfies $E(\varepsilon_t)=0$ and $Var(\varepsilon_t)=\sigma^2>0$.

A series that satisfies the above distribution is a smooth series, but in practical problems, many series have trends and are not smooth, which requires differencing the series to make it smooth. If the series contains seasonality, it is also necessary to perform seasonal differencing according to the seasonal cycle. At this point the $ARIMA(p,q)$ model can be improved to $ARIMA(p,d,q)(P,D,Q)s$. Where p is the autoregressive order, d is the smoothed difference score, and q is the moving average order. p is the seasonal autoregressive order, d is the seasonal smoothed difference, q is the seasonal moving average, and s is the number of seasonal cycles.

4. Spatio-temporal data mining of cross-border e-commerce consumer behavior

In this chapter, a cross-border e-commerce platform in China is selected as the research object, and a spatio-temporal data mining model of consumer behavior is used to analyze its real user behavior dataset, obtaining spatio-temporal characteristic data in terms of user behavioral transformations, user shopping time intervals, and user repurchase. Then, the model is used to analyze the behavioral data by clustering to obtain the portraits of different customer consumer groups, and to give different marketing strategy suggestions by combining the characteristics of different consumer groups.

4.1. User Behavioral Characteristics

4.1.1. Overall user analysis. This subsection focuses on the overall analysis of user behavior data, observing the trends of the four types of user behavior: browsing, collecting, adding purchase, and purchasing, as well as the differences and similarities of different behaviors before and after the promotional activities. The overall analysis of user behavior can discover the consumption habits and consumption tendency of users at different time nodes, predict the level of interaction before and after the activity period, and thus tap potential consumers, thus increasing product sales and store sales. In order to stimulate users' consumption and increase merchants' sales during various e-commerce festivals mainly in the form of promotional activities, e-commerce platforms will organize various interactive activities to guide consumers to interact directly with brands or products, resulting in the data volume of various behaviors during the event period being significantly higher than that during the flat sales period. In order to clearly observe the trend of the four behaviors, this paper selects the e-commerce festival with the event date of June 18th (“618”), and calculates the proportion of different behaviors every day during the festival period in June, and the result is shown in Figure 3.

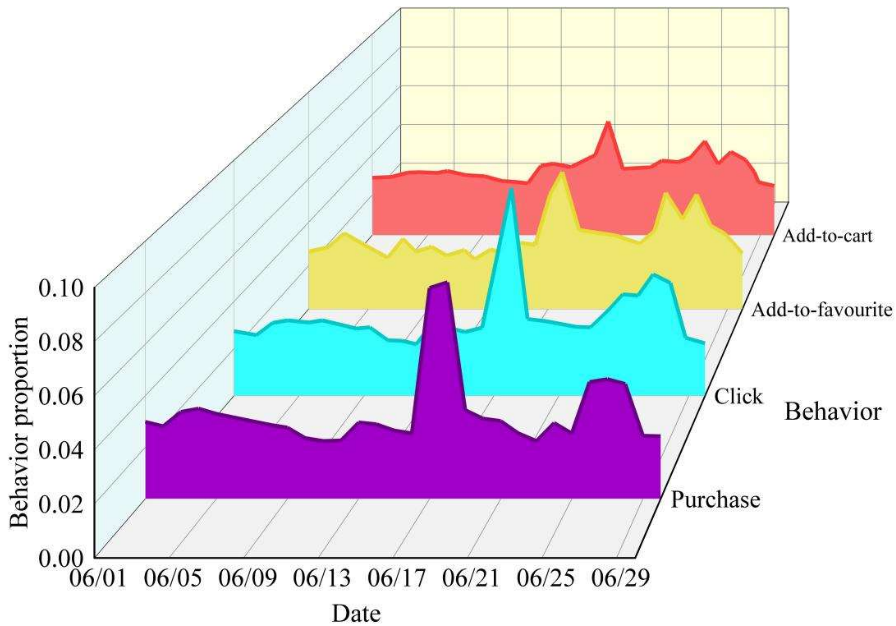


Figure 3. The trend of different behaviors over time

From Figure 3, it can be found that the trends of the four behaviors over time are similar, i.e., the percentage of user behaviors in the active period is significantly higher than that in the flat sales period. During the "618" activity, the proportion of the four behaviors is significantly higher than that of the flat sales period, among which the purchase behavior is the most prominent, which reflects the "big promotion effect" of the merchant's activities and the "promotion mentality" of the user, that is, the user browses, collects, and purchases more goods during the flat sales period, which is mainly prepared for the formal event, and the user directly buys after the activity starts, which leads to the outbreak of purchase behavior during the activity period. In addition, it can also be found that between June 25 and June 28, there was a second peak in the percentage of various behaviors, and there are three reasons for the analysis:

1) Merchants' secondary promotional activities. Merchants took advantage of the afterglow of the event to conduct secondary promotional activities to attract users to continue shopping on the platform, which led to a significant increase in various behaviors.

2) Influenced by the price comparison, return, re-purchase, and checking logistics information behaviors of users who have already purchased. For users who have already purchased during the activity period, users may have a price comparison mentality, and after receiving the goods, users may check whether there is a price reduction or price increase of the goods they purchased during the “618” activity period, which is also a major factor affecting the increase in the number of views. Commodities, users will return and secondary purchase, so a variety of behavior is bound to increase. In addition, due to the significant increase in sales of merchants during the activity period, the number of store shipments is high, the logistics pressure leads to express backlog, some of the goods can not be delivered to the user in a timely manner, the user is bound to query the logistics information, which will also make the user's behavior increase.

3) Influenced by the cross-border e-commerce platform set in July e-commerce festival. In July, major brands will hold relevant promotional activities, as an important activity to promote the brand, merchants will invest a lot of money in this activity, and before this put a lot of advertising for the activity of water storage, which will lead to an increase in user behavior.

4.1.2. Time interval analysis of user behavior. Each user's shopping behavior reflects the user's psychological changes to a certain extent, and deeper excavation of the information behind the user's behavioral changes can better grasp the user's purchasing trends. Therefore, this paper analyzes the time interval between the conversion of browsing, collecting, and adding shopping cart to purchase. When different behaviors are converted to purchase the longer the time interval, it means that consumers are more and more hesitant, they may be in the comparison from a large number of products so as to select the most preferred product, the longer the hesitation time, the smaller the possibility of purchase.

The proportion of users from browsing to purchasing at different time intervals is shown in Figure 4, where the X-axis indicates the number of days between users' conversion from browsing to purchasing, and the Y-axis indicates the proportion of the total number of behaviors that conform to the number of days between such intervals.

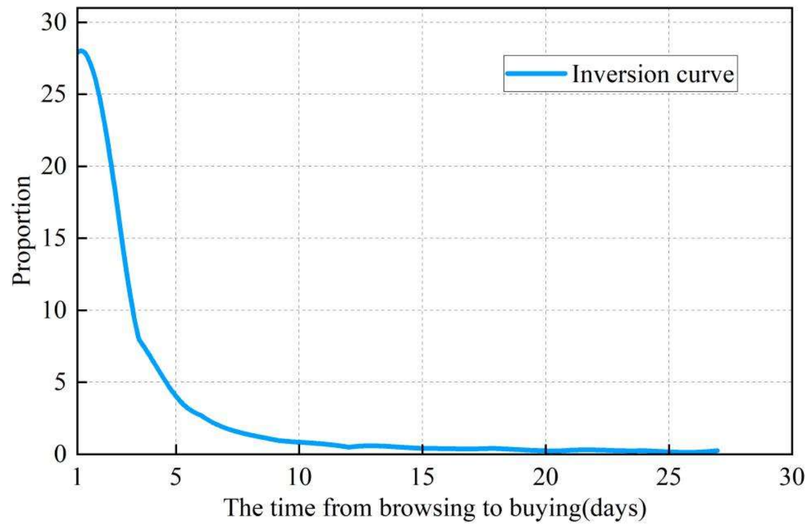


Figure 4. Time interval from browsing to purchase

From the figure, we can find that the time interval from browsing to purchase is mostly within 15 days, that is, users usually browse the goods within 15 days will buy the goods. Therefore, when merchants do advertising, you can focus on the last 15 days have browsed the product behavior of the user advertising, so that not only to determine the product purchase target group, but also effectively improve the conversion rate of product purchase.

Figure 5 for the user conversion from collection to purchase of different time intervals users accounted for the situation, in which the X-axis is the collection of conversion to purchase the interval of days, the Y-axis is in line with the number of days of this interval of the number of behaviors accounted for the proportion of the total number of behaviors.

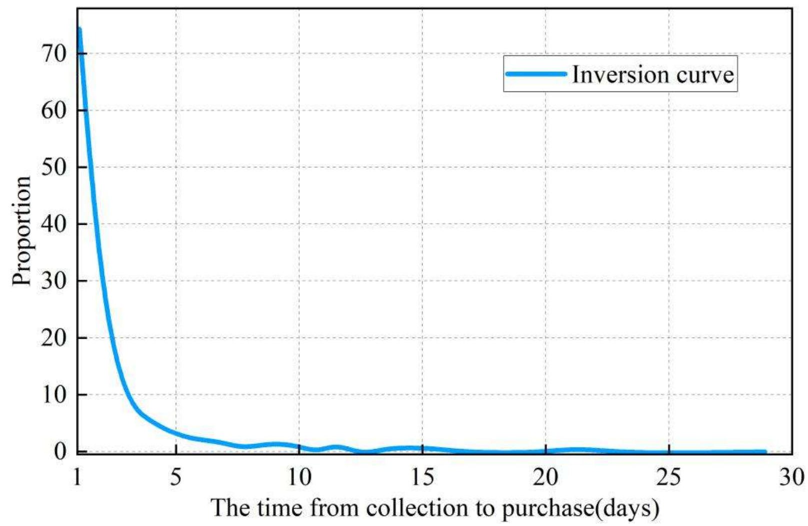


Figure 5. The interval between collection and purchase

From the figure, it can be found that most of the time intervals between conversion from favorites to purchases are less than 10 days, i.e., users usually buy this product within 10 days of favoriting the product. The proportion of purchase behavior occurring within nearly 1 day after collection reaches more than 70%, that is, users will buy the product in time after collection. At the same time, from the above figure can be seen, some users will be in the collection of goods 20 days after the purchase behavior, this may be because the user collection of goods after forgetting to pay, and did not buy other similar products, therefore, for the collection of products but long time did not buy the user, merchants can be appropriate to "remind" the user to let it pay.

Figure 6 for the user from adding a shopping cart into the purchase of different time intervals between users accounted for the situation, where the X-axis is to add a shopping cart into the purchase of the interval of days, the Y-axis is in line with the number of days between the number of behaviors accounted for the total number of rows of the proportion of the number of behaviors.

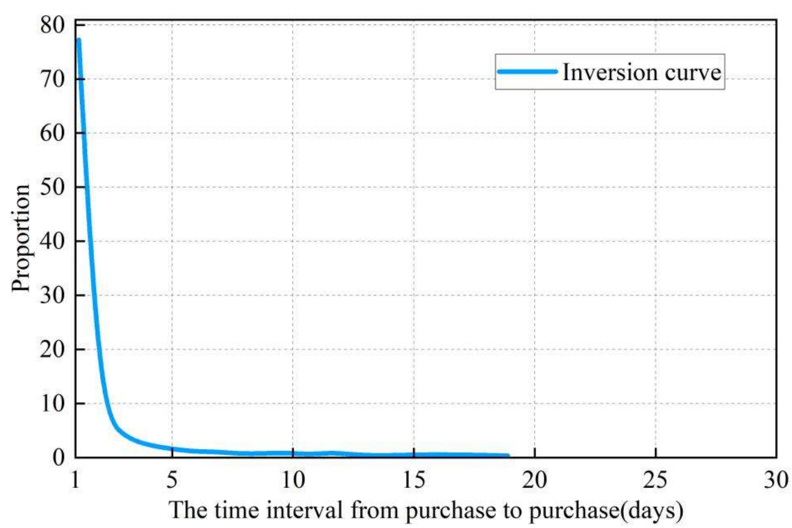


Figure 6. The time interval from purchase to purchase

It can be found that the majority of days between adding a shopping cart into the purchase of most of the days within 5 days, that is, the user is usually added to the purchase of goods within 5 days will buy this product, and in the shopping cart within the last 1 day of the purchase behavior accounted for more than 75%, that is, the user will be added to the purchase in time to buy the product. At the same time, it is not difficult to find that the above chart there is an obvious trailing phenomenon, the reason may be that the user to join the shopping cart in waiting for the price reduction of the product, or part of the user purely because of the like of certain products and add it to the shopping cart, but will not product to buy, and even more so the shopping cart of the goods as their own wish list. Therefore, for users who have added to the shopping cart behavior, the store can recommend the best combination of matching goods according to the user's interests and preferences, so that the user has the best experience of using the goods at the same time can also enhance the sales of store products. For users who have added to the shopping cart but have not paid for a long time, the store can remind the user to buy the product in time when the price drops or recommend similar products with low price and high cost performance in the store, so as to improve the overall sales of the store.

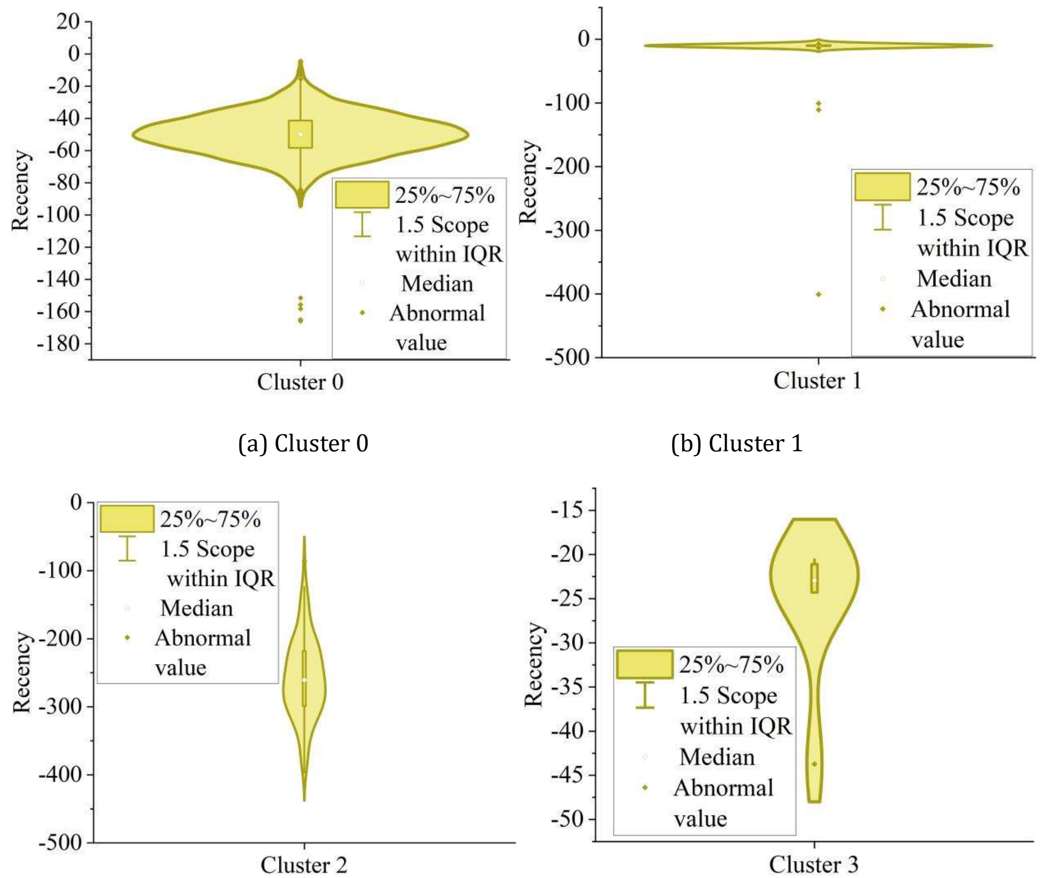
4.2. Consumer Profiling and Analysis

According to the results of the above analysis, taking into account the cross-border e-commerce platform sales management and computational efficiency and other issues, the number of clustering clusters will finally take the value of 4. The clustering center characteristics data of the four customer groups are shown in Table 1.

Table 1. Cluster analysis feature

Client group	Recency(R)	Frequency(F)	Monetary(M)
Cluster 0	0.5020188	-0.03215449	-0.04908449
Cluster 1	0.81159178	3.35362245	1.90541692
Cluster 2	-1.5407525	-0.3481604	-0.18526409
Cluster 3	0.85902946	12.4115325	18.90401738

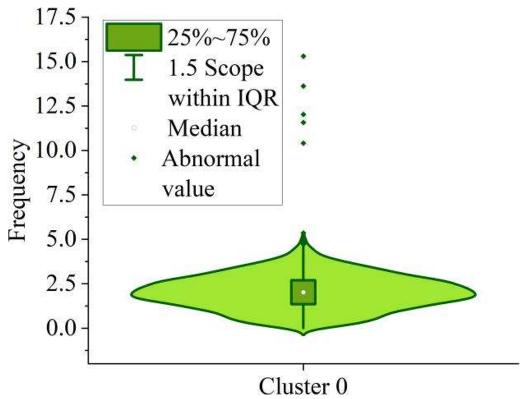
The clustering algorithm can only help to classify the customer group, but the specific group is what attributes of the customers need to be analyzed again, according to the clustering results of the distribution of customer group characteristics of the data drawn to help customer group analysis, the distribution of customer groups under the R dimension is shown in Fig. 7, the distribution of customer groups under the F dimension is shown in Fig. 8, and the distribution of customer groups under the M dimension is shown in Fig. 9.



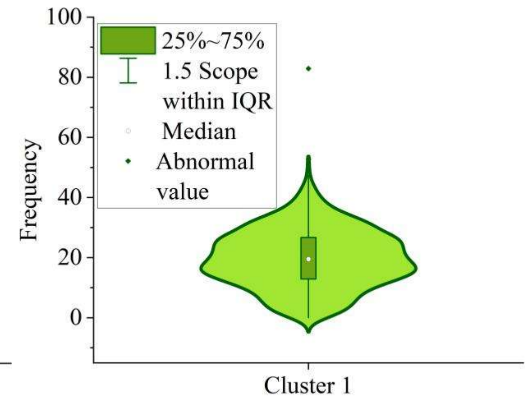
(a) Cluster 0 (b) Cluster 1

(c) Cluster 2 (d) Cluster 3

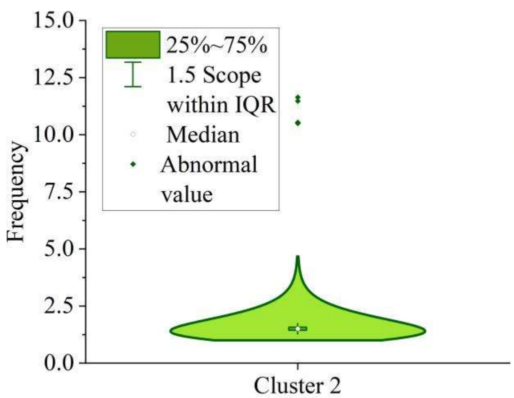
Figure 7. Analysis of clustering customers' Recency



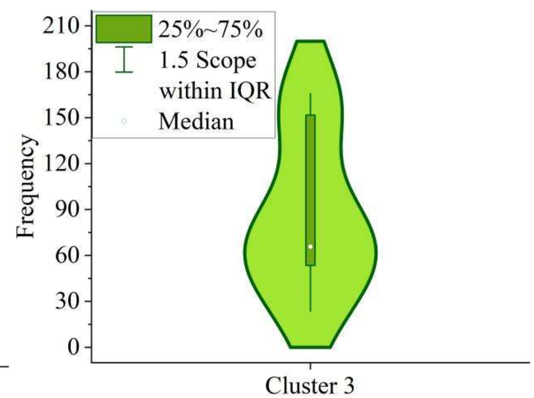
(a) Cluster 0



(b) Cluster 1

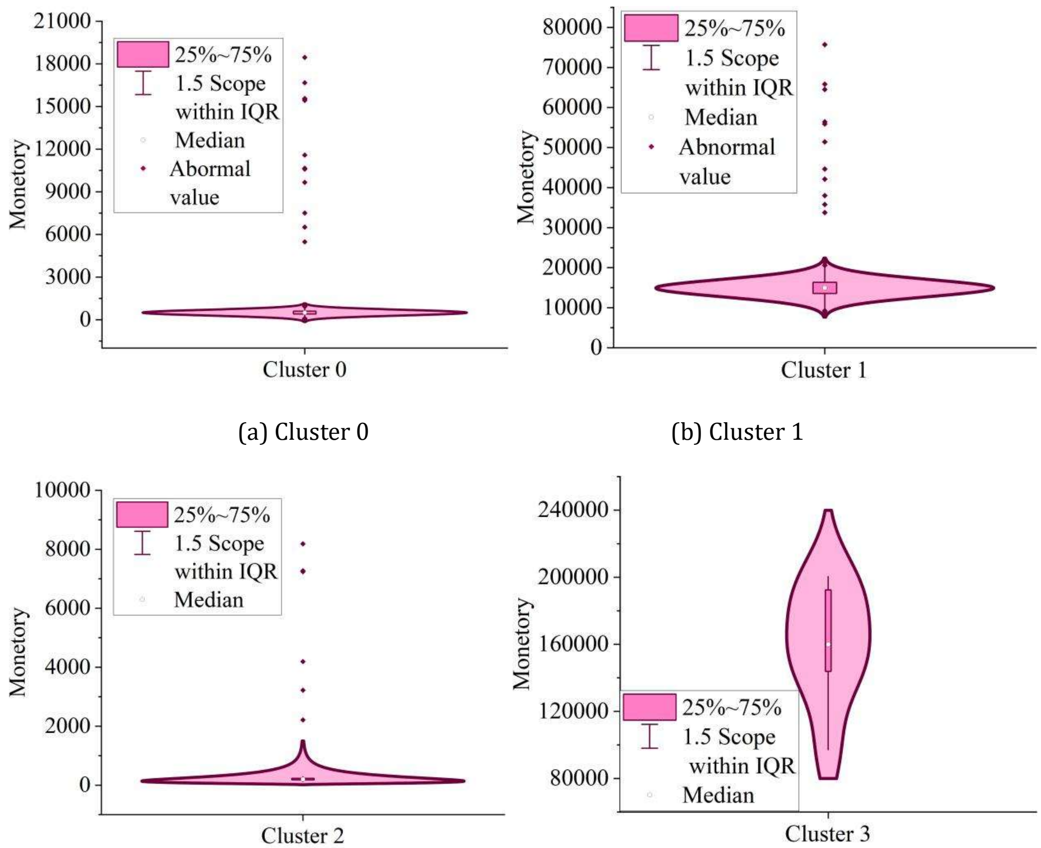


(c) Cluster 2



(d) Cluster 3

Figure 8. Analysis of clustering customers' Frequency



(a) Cluster 0 (b) Cluster 1

(c) Cluster 2 (d) Cluster 3

Figure 9. Analysis of clustering customers' Monetory

Observation of Figures 7-9 shows that:

1) The customer group of category 0, R, has a low M value and a relatively high F value, and although there is no consumption behavior in the last week, the frequency of past consumption is within 2-5 times. It belongs to the general retention customers, this customer group can be retained without spending too much investment by sending notices in promotions and giving some coupons.

2) The customer group of category 1 has higher R value, lower F & M, and has had consumption behavior within a week. It belongs to the customers who have recently consumed (likely to be new customers), and this customer group may contain many potential customers, which need to send out frequent campaign messages to increase the period consumption frequency and consumption amount.

3) Category 2 customer groups R, F, M are low, for the three low type, belong to the need to activate the customer, the customer group need to pay some attention, through frequent sending activity information, give some coupons and other forms of re-activation of consumption.

4) Category 3 customer groups R, F and M are higher, not only in a year with more than 50 times of consumption behavior, and the single consumption amount is higher. It is an important value customer, which needs to be focused on and given special attention services.

By carrying out visualization can intuitively feel the clustering results, so this paper also draws a scatter plot of the clustering results on different dimensional features see Figure 10.

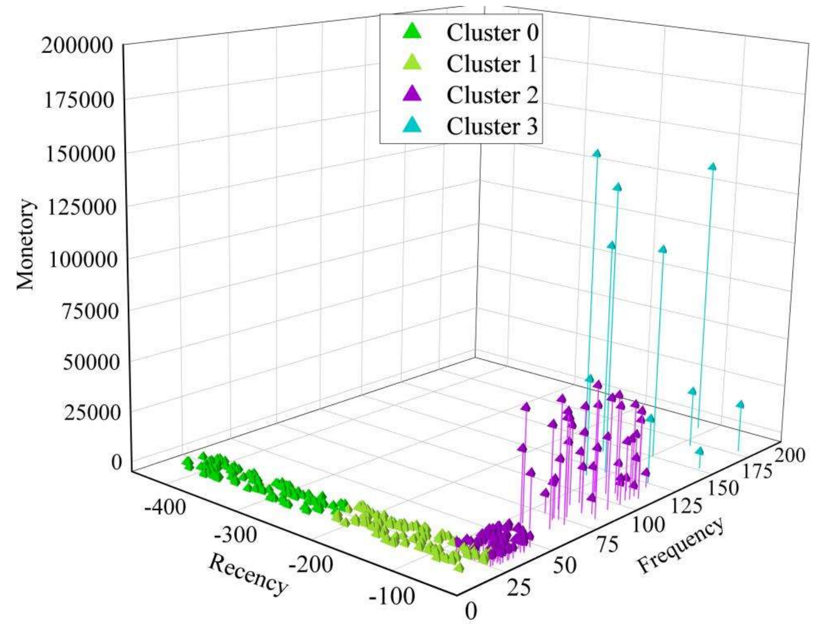


Figure 10. Customer RFM model clustering scatter chart

As can be seen from the figure, customer group 1 customers and other categories of customers are more differentiated, belonging to the three low customers. Customer group 0 has the largest proportion of customers, and the main difference with customer group 1 is that the recent consumption interval and the number of better, higher quality customers are customer group 2, generally have a recent consumption situation and consumption amount and consumption frequency of larger rise. The opposite of customer group 1 is customer group 3, which is very small in number and very high in quality. It can be seen that customers in the RFM three-dimensional features show regional distribution, which indicates that the clustering results can be well divided into different types of customers, so that cross-border e-commerce enterprises can formulate different marketing strategies and promotional activities to target different customer groups.

5. Conclusion

Based on the platform user behavior data, this paper takes the improved K-Means clustering algorithm as the core algorithm to realize the integration of clustering analysis technology and cross-border e-commerce platform.

In the data mining algorithm, for the defect that the K-Means algorithm does not take into account the inaccurate expression of clustering information caused by the differences in sample characteristics, the entropy theory knowledge is used to optimize the weights and achieve an objective and rational description of the characteristics of the data itself. And based on the improved clustering algorithm, the spatio-temporal data mining model of consumer behavior is designed. In the spatio-temporal data mining of cross-border e-commerce consumer behavior, the spatio-temporal data mining model of consumer behavior is used to draw four types of consumer portraits, namely, general retention, recent consumption, need to activate, and important value, and then different operational suggestions are given by combining the characteristics of different portraits.

References

[1] Chen, J., Lan, Y. C., & Chang, Y. W. (2023). Consumer behaviour in cross‐border e‐commerce: Systematic literature review and future research agenda. International Journal of Consumer Studies,

47(6), 2609-2669.

[2] Xiao, L., Guo, F., Yu, F., & Liu, S. (2019). The effects of online shopping context cues on consumers' purchase intention for cross-border E-Commerce sustainability. Sustainability, 11(10), 2777.

[3] Yang, Y., Yang, L., Chen, H., Yang, J., & Fan, C. (2020). Risk factors of consumer switching behaviour for cross-border e-commerce mobile platform. International Journal of Mobile Communications, 18(6), 641-664.

[4] Huang, S. L., & Chang, Y. C. (2019). Cross-border e-commerce: consumers' intention to shop on foreign websites. Internet Research, 29(6), 1256-1279.

[5] Valarezo, Á., Pérez-Amaral, T., Garín-Muñoz, T., García, I. H., & López, R. (2018). Drivers and barriers to cross-border e-commerce: Evidence from Spanish individual behavior. Telecommunications Policy, 42(6), 464-473.

[6] Li, Y., & Deng, Y. (2023). Analysis of the impact of cross-border e-commerce on consumer behavior anxiety from the perspective of consumer psychology. CNS Spectrums, 28(S1), S31-S32.

[7] Wang, C., Liu, T., Zhu, Y., Wang, H., Wang, X., & Zhao, S. (2023). The influence of consumer perception on purchase intention: Evidence from cross-border E-commerce platforms. Heliyon, 9(11).

[8] He, W., & Wisaeng, K. (2024). UTILIZING CROSS-BORDER E-COMMERCE TO PROMOTE ECONOMIC SUSTAINABLE DEVELOPMENT: LESSONS LEARNED FROM CHINA'S TRIAL ZONES. International Journal of eBusiness and eGovernment Studies, 16(3), 137-161.

[9] Yang, Y., & Lin, W. L. (2022). The Impact of Consumer Trust and Consumer Loyalty on Sustainable Development of Cross-border E-commerce. Special Education, 1(43).

[10] Guo, Z., & Chelliah, S. (2024). What influences cross-border e-commerce consumer satisfaction? A global consumer perspective. Global Business & Management Research, 16.

[11] Lin, A. J., Li, E. Y., & Lee, S. Y. (2018). Dysfunctional customer behavior in cross-border e-commerce: A justice-affect-behavior model. Journal of Electronic Commerce Research, 19(1), 36-54.

[12] Lu, H., & Liang, J. (2024, August). Consumer Behavior Analysis Model of Cross-border E-commerce Based on Big Data. In 2024 International Conference on Power, Electrical Engineering, Electronics and Control (PEEEC) (pp. 722-726). IEEE.

[13] Vijayalakshmi, S., Shanmugasundaram, S., Padmanabhan, P., & Jerald Nirmal Kumar, S. (2024). Spatio-Temporal Supply Chains and E-Commerce. In Spatiotemporal Data Analytics and Modeling: Techniques and Applications (pp. 179-192). Singapore: Springer Nature Singapore.

[14] Mu, W. (2019). A Big Data-based Prediction Model for Purchase Decisions of Consumers on Cross-border Ecommerce Platforms. Journal Européen des Systèmes Automatisés, 52(4).