J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 1117--1131

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

Research on the influence of nonlinear regression model of user data on sales strategy adjustment in cross-border e-commerce Wish platform

Dan Rong[1,✉]

[1] Business School, Ningbo City College of Vocational Technology, Ningbo, Zhejiang, 315211, China

ABSTRACT

Under the accelerated process of economic globalization and the booming development of Internet technology, cross-border e-commerce, as a new mode of international trade, is becoming a new driving force for the transformation and upgrading of foreign trade with its high efficiency and convenience, low cost and high benefit. This study uses data cleaning and missing value filling methods to preprocess user behavior data and merchandise sales marketing data in cross-border e-commerce Wish platform, and discretizes user behavior data using rough set method. Then, we select the merchandise sales and user behavior as the dependent and independent variables to construct a multiple nonlinear regression model in order to analyze the influence of user data on sales in cross-border e-commerce Wish platform. The results of the multivariate nonlinear regression model show that user behavior in cross-border e-commerce Wish platform has a significant effect on merchandise sales (P=0.005243). It is also found that the sales strategy adjusted according to the regression results can improve the sales and promotion effect of enterprises in cross-border e-commerce platform. The research results of this paper enrich the theoretical and practical research on the optimization and adjustment of cross-border e-commerce enterprises' sales strategies, provide theoretical basis and decision-making reference for the subsequent adjustment of cross-border e-commerce enterprises' sales strategies, and help cross-border e-commerce enterprises to go global.

Keywords: multiple nonlinear regression, rough set, cross-border e-commerce, sales strategy

1. Introduction

With the acceleration of globalization, cross-border e-commerce has become a business trend that cannot be ignored. More and more consumers start to buy goods overseas to get higher quality and lower price. And wish cross-border e-commerce platform is one of the most popular ones. Wish cross-border e-commerce platform is a global e-commerce platform that provides various types of goods

✉ Corresponding author.
E-mail address: Rdv1025@163.com (D. Rong).
Received 05 January 2024; Revised 20 May 2024; Accepted 15 October 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-062
© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

and services to consumers around the world. The main features of the platform are low prices and a wide range of products [1-2]. It provides a convenient way for consumers to purchase goods from all over the world. Wish platform has a wide range of goods, including clothing, shoes, household goods, digital products, beauty products, toys and other areas. The prices of these products are low because they come from small producers around the world. These producers usually don't have traditional sales channels, so they can market their goods through the Wish platform, a globally recognized e-commerce platform that attracts a large number of sellers and buyers. However, to be successful on Wish, some effective operational ideas and strategies are needed [3-6].

On the Wish platform, data analysis is the key to developing effective promotion strategies. Through in-depth analysis of Wish platform user data, sellers can more accurately understand the market demand, develop targeted promotional strategies to enhance the marketing effect and achieve sustained business growth. From keyword optimization, advertising, social media marketing to shopping experience optimization and other aspects, data analysis plays a crucial role [7-8].

Literature [9] aims to analyze and identify effective digital marketing strategies to increase sales conversion on e-commerce platforms. Emphasis was placed on data analysis by recording important information through data collection techniques. The results show that digital marketing strategies can play an important role in improving sales conversion rates on e-commerce platforms. Literature [10] examined the factors influencing online sales on O2O on-demand versus traditional B2C platforms. Based on sample data from e-commerce platforms, it was emphasized that there are differences in the influence mechanisms and the degree of influence of each factor between O2O and B2C platforms. Literature [11] explored the application of big data analytics in e-commerce marketing strategies, emphasizing the use of customer data to provide volume personalized experiences and recommendations. Based on case studies and industry practices, it reveals that big data-driven personalization facilitates customer engagement and loyalty. Literature [12] aims to identify research trends in this area through LRSB of e-commerce marketing strategy research. The database-based literature review indicated that companies respond to e-commerce and online business with the help of e-commerce platforms and social networks, which drive marketing strategies and share information on innovations by understanding consumer needs. Literature [13] aims to analyze the impact of digital marketing on product quality and organizational competitiveness. It is pointed out that digital marketing helps to target content that is personalized to the needs of consumers, which not only provides companies with the opportunity to interact with their audience, but also helps to help them improve product quality, which is essential for their success in a dynamic competitive environment. Literature [14] explores the application of recommendation algorithms in e-commerce platforms and analyzes the advantages and disadvantages of algorithms such as collaborative filtering and hybrid recommendation models and the overall impact of these algorithms on the platforms through personalized recommendations. Based on case studies, it reveals that recommendation algorithms play an important role in the user shopping experience and the development of e-commerce platforms. Literature [15] derives the characteristics of user types based on the collection and analysis of data such as user characteristics and consumption behavior. Precision marketing strategies for different groups are proposed, and the implementation effect is evaluated. The results proved the effectiveness of the precision marketing strategy. Literature [16] aims to create a mechanism for designing a precision marketing planning platform based on user data mining. And the platform including user profiling system and so on was designed. By helping e-commerce platforms to implement precise marketing and advertising strategies in order to realize the business value in the context of big data. Literature [17] established a user value model based on user data. The improved K-means algorithm is used to cluster users' purchasing behavior, construct customer value matrix, and classify users. It is verified through testing that the analysis of user clustering features by the improved K-means algorithm is conducive to the development of more accurate marketing strategies. Literature [18] emphasizes the significant impact of big data on business models and enterprise development. And pointed out that in the current era, enterprises should pay attention to the value mining of data information, so as to realize accurate marketing. Literature [19] reveals the important role of precision marketing of big data information, and introduces that enterprises are able to achieve personalized marketing strategies based on data mining and analysis to obtain user information. It also analyzes the precision marketing of e-commerce enterprises and puts forward corresponding marketing strategies.

This paper collects e-commerce product marketing data and user-related behavior data from the cross-border e-commerce Wish platform, and first evaluates the quality of the data and directly removes unreasonable data from the data. Using direct deletion or missing value filling and other ways to deal with missing data, and associated cleaning and integration of user behavior and product marketing data. Then the rough set method is applied to process the user behavior data, and the attribute values of the user data are discretized, so that they are transformed into stereotyped data, which facilitates the analysis of the multivariate nonlinear regression model. Subsequently, according to the purpose of the study, the marketing amount of goods in the cross-border e-commerce Wish platform is determined as the dependent variable, and 10 variables, such as user gender, membership level, and average monthly consumption, are taken as independent variables, and the multivariate nonlinear regression model is constructed to explore the degree of influence of user behavior on the sales of goods. Company A is selected as a research case, and based on the company's user behavior and marketing data in the cross-border e-commerce Wish platform for multiple nonlinear regression analysis, and according to the results of the regression analysis to adjust the optimization of the sales strategy, to explore the implementation of the optimization strategy after the implementation of the sales of the product to improve the situation.

2. Cross-border e-commerce platform user dataset construction

The research data in this paper comes from the e-commerce marketing data available on the cross-border e-commerce Wish data analytics platform, the corresponding product display pages, and the relevant sales product information data in specialized product information websites. It contains the statistical data of the products being sold on the cross-border e-commerce Wish platform within the statistical time period, the data of the product's display page and the corresponding product information, as well as the product ratings. Before establishing the user nonlinear multiple regression analysis model, it is first necessary to deal with a large number of original tables and original fields, and carry out a large number of data preprocessing work, such as the processing of missing values and outliers, and correlating multiple source tables to get the target fields, and so on. And if directly in the cross-border e-commerce Wish platform database or data warehouse to go to these data preprocessing, and direct data modeling work, obviously is not reasonable. Therefore, according to the needs of the data mining project, we need to organize a separate data table or several data tables specifically for data mining, these data tables are obtained through the source database or data warehouse source table association and merger processing operations, these data tables are the subsequent nonlinear regression modeling modeling of the basic data set.

2.1. Data collection and selection

Before the data mining work is carried out, it is necessary to first determine where the source of the data is, collect the data in advance and store it, and what kind of data can be obtained through these data. When getting these raw data, the first thing to do is to analyze and study these data, and according to the goal of data mining, choose the suitable raw data and organize them into a number of tables to form a data set.

2.2. Data quality assessment and integration

2.2.1. Data quality assessment. After forming the cross-border e-commerce Wish platform user data set, the next step is to assess and clean and organize the quality of the data to ensure that the quality of the data provided to the data mining work is good, so as to ensure the effectiveness of the data model. For data values from different data sources, it is important to ensure that the statistical caliber should be consistent. When analyzing the quality of data, usually encounter a variety of unreasonable data problems, for example, the age of the user is more than two hundred years old or negative values, which is obviously unreasonable, there must be data statistical problems. When integrating different data sources, it is necessary to pay attention to the unit or statistical caliber of the same field in different data sources to maintain consistency.

2.2.2. Missing data handling methods. The problem of missing data should also be taken very seriously, because missing data can seriously affect the subsequent modeling work. If the number of samples with missing values is much smaller than the total number of samples, this sample with missing values can be deleted directly in order to facilitate the process. However, when the number of samples with missing values is large, if the samples with missing values are deleted directly, the number of samples left at the end will be very small and a lot of information will be lost. At this time, the method of missing value filling [20] can be taken to find the alternative values of missing values. For numerical variables more commonly used missing value filling method is to use the middle value, the average value to fill, or according to the distribution of the overall data to fill the data. It is also possible to predict the missing values by building a predictive model of this missing value variable, but this practice pays a high cost.

2.2.3. User data integration. As the source data used for modeling the nonlinear regression analysis model of cross-border e-commerce Wish platform users are likely to come from different databases or external data sources, and the distribution is more fragmented, it is necessary to merge and integrate the data. Through data correlation and cleaning, the data from different data sources will be integrated into one or several tables, and it should be noted that in the process of collation, it is necessary to ensure that the caliber of statistics between fields should be consistent. If the caliber of the data is not uniform, it will make the quality of the mined data worse. Different data sources of inconsistent statistical caliber of the solution is relatively simple, such as some of the amount of units are used in dollars, while some amount of units are used in cents, then as long as the unit change can be resolved, and for example, for the same user, some of the table with the user's login name, some of them with the user's name, which only needs to be associated with a user, unified for a fixed user name.

2.3. Discretization of user information attribute values

Data can be divided into qualitative data and quantitative data. For qualitative data, the data itself is a numerical value, but the meaning it represents does not necessarily have numerical significance, for example, "1" and "0" can be used to indicate the gender of the person respectively, if the designation of "1" means "male", then "0" means "female", here the form of the number "1" and "0" are essentially two symbols, used to distinguish the gender of the person, in this sense, you can also specify the other two values, to achieve the same effect.

The classical rough set method [21] is more suitable for dealing with data whose attribute values are qualitative, which is also known as symbolic data, and the domain of attribute values in information systems whose attribute values are qualitative is generally a finite set, i.e., there are only a finite number of values for each attribute of all objects. Most of the attribute values of the objects in the data covered in this paper are real-type data, and the meaning is the numerical meaning expressed by the values themselves, such a system is called a continuous-valued information system in rough set theory.

In order to facilitate processing by rough set methods, the user data attribute value domain needs to be discretized. That is, the quantitative data into qualitative data, the original attribute value domain is a real number interval, after the transformation of the value domain is a finite set, the elements of which may be the form of the data is still the same, may also be a string to represent the symbols.

Here for the discretization of attribute value domains in a continuous-valued information system, the approach of equidistributing each value domain interval is used. Specifically, let (U, A, AT, f) be a continuous-valued information system, and for any $a \in A$, notation:

$$m_a = \min\left\{f(x,a) \mid x \in U\right\}, M_a = \max\left\{f(x,a) \mid x \in U\right\} \tag{1}$$

Reset the value field of attribute a to $V_a = [m_a, M_a]$. Divide interval V_a into k subintervals of equal length, so that each subinterval has length equal to $d_a = \dfrac{M_a - m_a}{k}$ and the $i(1 \le i \le k)$ th subinterval is.

$$\left[L_a^i, R_a^i\right] = \left[m_a + (i-1)d_a, m_a + id_a\right] \tag{2}$$

Or:

$$\left[L_a^i, R_a^i\right] = \left[M_a - (k-i)d_a, M_a - (k-i+1)d_a\right] \tag{3}$$

The process of discretizing the value domain of attribute a corresponds to a mapping T_a from the original value domain V_a to the new value domain $V_a^* = \left\{v_a^1, v_a^2, \cdots, v_a^k\right\}$, for any $v \in V_a$, when $v \in \left[L_a^i, R_a^i\right]$, or $v = R_a^i = L_a^{i+1}$, is defined $T_a(v) = i$.

By discretizing the value domain of each attribute, the original continuous value information system is transformed into a symbolic value information system.

3. User nonlinear regression modeling methods

3.1. Multiple nonlinear regression models

In the actual problem of the influence of cross-border e-commerce Wish platform users on the adjustment of sales strategy, there are often more than one factor affecting the dependent variable, and this kind of regression problem with one dependent variable and multiple independent variables is called a multiple regression problem. When the dependent variable and each white variable is a linear relationship, it is called multiple linear regression, and when the dependent variable and each independent variable is a nonlinear relationship, it is called multiple nonlinear regression [22]. The principle of multivariate nonlinear regression is basically the same as that of univariate nonlinear regression, but it is much more complicated in calculation, so it is mostly done with the help of computer software.

Let the dependent variable be y and each of the k independent variables be $x_1, x_2 \ldots\ldots x_k$. The equation describing how dependent variable y depends on independent variable $x_1, x_2 \ldots\ldots x_k$ and error term ε becomes a multiple regression model, which can be expressed in general form as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k + \varepsilon \tag{4}$$

where ε is the error term. y is the linear function $\beta_n + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k$ of $x_1, x_2 \ldots\ldots x_k$ plus the error term ε. The error term reflects the effect on y of random factors other than the linear relationship between $x_1, x_2 \ldots\ldots x_k$ and y, and is variability that cannot be explained by the linear relationship between $x_1, x_2 \ldots\ldots x_k$ and y.

In a multiple nonlinear regression model, the same three basic assumptions are made about the error term ε.

1) The error term ε is a random variable with expectation 0, i.e., $E(\varepsilon)^{-0}$. This implies that for a given value of $x_1, x_2 \ldots\ldots x_k$, the expectation of y is $F(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k$.

2) The variance δ^2 of ε is the same for all values of the independent variable $x_1, x_2.....x_k$.
3) The error term ε is a random variable that obeys a normal distribution and is independent of each other, i.e., there is $\varepsilon \sim N(0,\sigma^2)$. Independence implies that ε corresponding to a particular set of values for the dependent variable $x_1, x_2....x_k$ is uncorrelated with ε corresponding to any other set of values for $x_1, x_2....x_k$. Normality implies that for a given value of $x_1, x_2 ......x_k$, the dependent variable y is also a random variable obeying a normal distribution.

According to the regression model it is assumed that there is:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ...\beta_k x_k \tag{5}$$

The above equation is called the multiple regression equation which describes the relationship between the expected value of the dependent variable y and the independent variable $x_1, x_2......x_k$.

Parameter $\beta_i \beta_2 ...\beta_k$ of the regression equation is unknown and sample data is used to estimate them. When the sample statistic is used to estimate the unknown parameter $\beta_1 \beta_2 ...\beta_k$ of the regression equation, the estimated multiple regression equation is obtained in its general form:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_k x_k \tag{6}$$

where $\hat{\beta}_0, \hat{\beta}_1 ...\hat{\beta}_k$ is the estimated value of parameter $\beta_1 \beta_2 ...\beta_k$, $\hat{y}$ is the estimated value of dependent variable y, where $\hat{\beta}_0, \hat{\beta}_1 \dots \hat{\beta}_k$ is called the partial regression coefficient, and $\hat{\beta}_1$ represents the average change per unit y change in x_1 when $x_1, x_2....x_k$ held constant. $\hat{\beta}_2$ denotes the average change per unit y change in x_2 when $x_1, x_2.....x_k$ constant, and the rest of the partial regression coefficients have similar meanings.

The $\hat{\beta}_0, \hat{\beta}_1 \dots \hat{\beta}_k$ in the regression equation is still based on the least squares method, that is, summing the squares of the residuals:

$$Q = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - .... - \hat{\beta}_k x_k)^2 \tag{7}$$

is minimized, from which the standard system of equations for solving $\hat{\beta}_0, \hat{\beta}_1 ...\hat{\beta}_k$ can be obtained as:

$$\begin{cases} \dfrac{\partial Q}{\partial \beta_0} |_{\beta = \hat{\beta}_0} = 0 \\ \dfrac{\partial Q}{\partial \beta_i} |_{\beta_i = \hat{\beta}_i} = 0 (i = 1, 2, ...k) \end{cases} \tag{8}$$

Solving the above system of equations can be done with the help of computer software and the results can be obtained directly using EXCEL software.

3.2. Regression model variable selection

This paper wants to study how the elements of user-related characteristics in the cross-border e-commerce Wish platform affect the e-commerce marketing effect of products in the cross-border e-commerce platform through a nonlinear multiple regression analysis model. That is, to find the existence of the influence elements in the adjustment of sales strategy, how these factors have an impact on the effect of product sales, and which factors of the adjustment of the effect of product sales is better. First of all, it is necessary to set and explain the target variables and condition variables.

3.2.1. Independent variable setting. Independent variables are variables used to explain the target variable (dependent variable), creating independent variables is the key to building an effective data model, and the appropriate choice of independent variables has a very important relationship with the effectiveness of the model. Before establishing the model, through analysis and research, it is found that the main factors that have an impact on user churn are the purchase of investment products in the last three months, the change of user assets in the last three months, the return of user payments and user attributes. In this paper, we will measure the impact on user churn from three aspects: user attributes, asset factors, and behavioral factors, and finally determine that there are 10 main variables affecting the churn of Internet financial users, and the abbreviations and explanatory notes of specific variables are shown in Table 1. Some of the indicators that represent the main behavior of the user, respectively, the average number of monthly product views in the last three months (MLD), the average monthly consumption in the last three months (MIA), the maximum amount of consumption in the last three months (MMIA), are related to the user's continuous behavior in the previous three months. The reason for the unit of days (a user may generate the same behavior several times in a day, but only recorded as 1) is to consider the continuity of user behavior, which may have an impact on user consumption.

Table 1. Cross-border e-commerce platform user variables

Number	Variable shortterm	Variable specification	Type	Variable interpretation
1	Sex	Gender	Varchar	Yes, no
2	Level	Membership grade	Varchar	1,2,3…,9
3	IFTI	Whether it is only the first consumption	Varchar	Yes, no
4	ISD	Whether there is a second consumption	Varchar	Yes, no
5	Age	Age	Int	——
6	IT	Consumption duration	Int	The first consumption is the current time (day)
7	MLD	Browsing frequency	Int	The number of visits for nearly three months
8	PS	Product satisfaction	Int	Product evaluation grade (1~5)
9	MIA	Monthly consumption	Money	Total consumption in nearly three months
10	MMIA	Maximum consumption	Money	Nearly three months of maximum consumption

3.2.2. Target variable setting. In the relevant research in marketing, the target variables of the empirical research on the effect of sales strategy implementation are mostly the actual sales data. According to the data source of this paper, the relevant information that can be obtained from the cross-border e-commerce Wish platform includes product sales volume and sales volume. The sales volume refers to the total number of products sold during the statistical period, and the sales volume can not only indicate the market share of the product, but also show the marketing ability of the product in the competition of similar products. Another data related to the implementation of the sales strategy is sales, which refers to the total revenue obtained from the sale of the product during the statistical period. For certain products with high added value, high sales volume does not at the same time mean higher sales. High sales represent that the product is recognized by consumers on the cross-border e-commerce Wish platform and they are willing to pay for it, but such high sales may be a direct result of thin profit margins, while the key that manufacturers obviously attach more importance to is whether the product can bring higher revenues for themselves. Therefore, this paper sets sales (XE) as the target variable.

4. Analysis of the impact of the user's nonlinear regression model on sales effectiveness

4.1. *Company A's cross-border platform stores*

Company A is located in Fujian Province, the main production area of tea, and with this geographical advantage, the company's core business is to operate tea trade, and in recent years, it gradually shifted its focus to the B2C trade of tea on e-commerce platforms.The development history of cross-border e-commerce of Company A is relatively short. Since January 2021, the company has cooperated with the cross-border e-commerce platform Wish and constructed its own store, thus initially implementing its cross-border e-commerce business. Its main online retail products cover tea and related merchandise, including multiple categories of tea, as well as tea sets and household products, Chinese herbs and herbal products, coffee and liquor, and Chinese specialty food products, etc. Company A hopes that the diverse range of merchandise will attract a wide range of customers to visit its store in the cross-border e-commerce Wish platform, and, due to the difficulties in obtaining export trade licenses, Company A hopes to maximize the use of such licenses to sell a Company A's main business model on the cross-border e-commerce Wish platform is retail. They adopt an online shipment model, where they package the ordered products themselves and deliver them to buyers through cross-border logistics.Company A's monthly sales of its store in the cross-border e-commerce Wish platform from January 2021- to December 2023 are shown in Figure 1.Company A's store design, merchandise design, and product design are all in accordance with the requirements of the Wish platform. Although Company A already has certain advantages in basic elements such as store design, product quality and logistics and distribution, their store belongs to the medium-level sellers according to the assessment of operational data, and the number of orders and payments is relatively low. According to the data, Company A's cross-border e-commerce Wish platform in the store average daily orders of about 14 single, the average monthly sales of 31,300 U.S. dollars. However, these figures are affected by seasonal factors, resulting in a fluctuating trend of slow growth in monthly sales. This may mean that Company A has more active sales during certain periods or seasonal events, while sales are relatively flat during other periods.

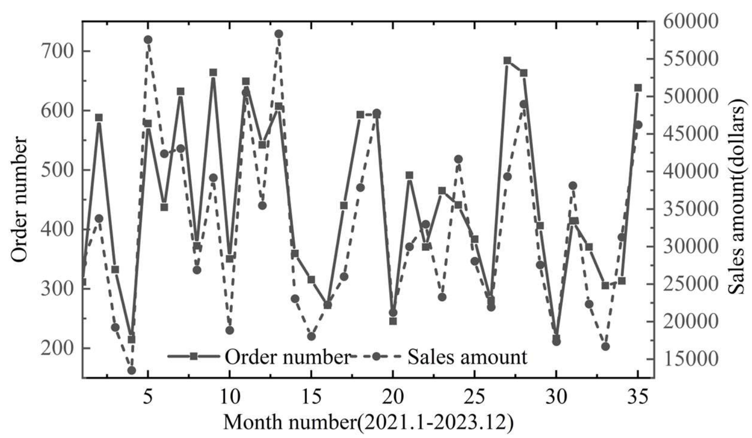

Fig. 1. The monthly sales of a company store in 2020-2023

Meanwhile, the actual sales data of Company A's e-commerce and the user composition-consumption times obtained from platform data analysis are shown in Figure 2, respectively. The main customer base of the store is independent consumers who make one-time purchases (25.49%), and its return rate is low. This has led to the store's sales relying mainly on retail sales, while the number of direct sales customers gained through the store is low, and the growth of sales is relatively slow, with no significant upward trend. Overall, the e-commerce trade in Company A's cross-border e-commerce Wish platform has begun to take shape, with sales gradually increasing and the sales market rapidly expanding to many countries and regions around the world. However, due to the lack of a complete set of sales strategy adjustment method, its cross-border sales performance from the platform head of tea e-commerce is still a big gap, the business level still has a lot of room for improvement.

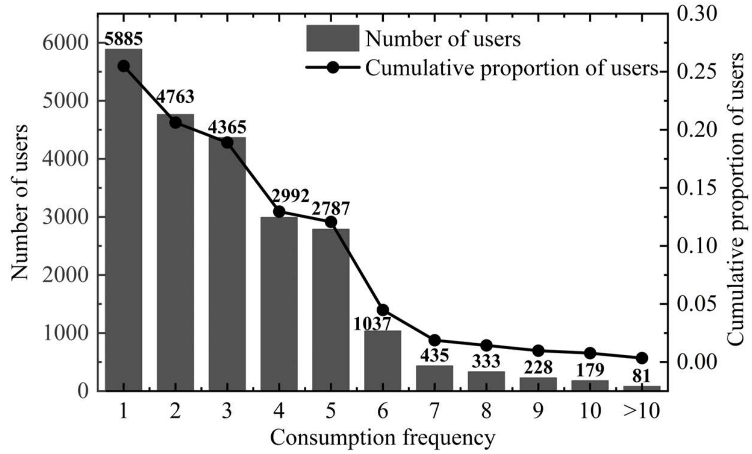

Fig. 2. User composition - consumption frequency analysis results

4.2. *Analysis of the impact of user characteristics on merchandising*

4.2.1. Variable correlation analysis. This paper constructs the user dataset of Company A in the cross-border e-commerce Wish platform based on the data collection and processing methods above. Before conducting multivariate nonlinear regression, the basic assumption of regression analysis needs to be satisfied, that is, there is a certain correlation between the variables. The results of the correlation analysis between the variables are shown in Figure 3, and it is found that the correlation coefficients between the variables are in the range of 0.212-0.683, none of which exceeds 0.7, which suggests that there is a correlation between the characteristic variables of the users in the cross-border e-commerce Wish platform and the sales of goods, and that a multivariate nonlinear regression analysis can be carried out.

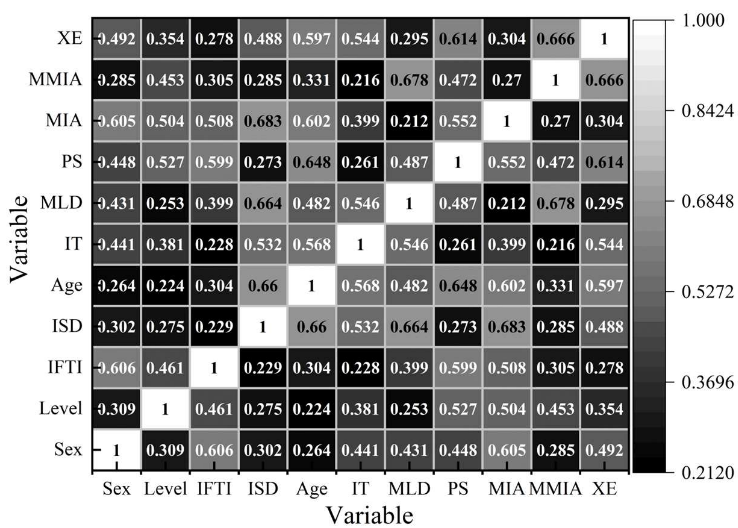

Fig. 3. Variable correlation analysis results

4.2.2. Regression model construction and testing. According to the data of user characteristics and sales in the cross-border e-commerce Wish platform of Company A, the multivariate nonlinear regression model of the user's each characteristic on product sales is constructed, and the results of the multivariate nonlinear regression analysis are shown in Table 2. Although the overall P-value of the model is small at 0.008452, it can pass the significance test of the equation, and the R-square is 0.9892, which indicates that the model fits well. However, looking at the P-value of each parameter individually, only 2 parameters (IFTI and IT) have P-values (0.0613 and 0.0595) less than 0.1, while the P-values of all other parameters are larger, suggesting that the regression coefficients do not pass the test of significance and that the model may be subject to multiple covariance. Covariance refers to the existence of very strong correlation between more than two explanatory variables, then these variables can actually be portrayed by the same variable, and the inclusion of these variables at the same time will lead to errors in the regression results, in order to avoid these problems, the variables need to be analyzed for correlation. Therefore, a stepwise regression analysis needs to be done on the model to eliminate the variables that do not have much impact on the results.

Table 2. Multivariate nonlinear regression analysis results

Independent variable	Estimate	Std.Error	t value	Pr(>|t|)
Sex	0.44299	0.9718	1.974	0.2163
Level	0.75139	0.9734	0.731	0.2496
IFTI	2.39416	0.6144	1.559	0.0613
ISD	1.72920	0.7511	1.717	0.2449
Age	0.18408	0.9132	0.627	0.2765
IT	2.27808	0.9066	0.523	0.0595
MLD	1.06526	0.6805	0.87	0.272
PS	1.74646	0.9829	0.41	0.1067
MIA	2.29664	0.887	1.733	0.2854
MMIA	1.40008	0.8148	1.303	0.2001
Multiple R-squared: 0.9975, Adjusted R-squared: 0.9892				
F-statistic: 115.26, p-value: 0.008452				

Stepwise regression analysis is to remove or add variables by choosing the smallest AIC information statistic as a guideline. In this paper, we use R language to do stepwise regression analysis, and the results of stepwise regression analysis are shown in Table 3. It can be seen that when the cross-border e-commerce Wish platform user's gender, age membership level, first time consumption, second time consumption, consumption length, product browsing times, average monthly consumption, maximum consumption and product satisfaction are used as the coefficients of the regression equation, the value of AIC is -91.52, and the value of the regression equation's AIC after the removal of the variable of the user's gender is -92.36. Since the removal of the cross-border e-commerce Wish platform, the AIC of the regression equation can be minimized by removing the variable of user's gender, so R automatically removes this item. After removing these two variables, the value of AIC of the regression equation will increase when removing any one of them, so the stepwise regression analysis is terminated and the optimal regression equation is obtained. It is found that the optimal multiple nonlinear regression equation can be obtained after removing the user gender variable in the cross-border e-commerce Wish platform.

Table 3. Stepwise regression analysis results

Start: AIC=-91.52				
XE~ Sex + Level + IFTI + ISD+ Age+ IT+ MLD+ PS+ MIA+ MMIA				
Independent variable	Df	Sum of Sq	RSS	AIC
Sex	1	0.00692	0.000897	-91.87
<none>	——	——	0.000546	-91.52

Level	1	0.006913	0.000566	-70.38
IFTI	1	0.004896	0.000639	-76.8
ISD	1	0.00605	0.000991	-72.86
Age	1	0.005977	0.000694	-74.29
IT	1	0.006423	0.000743	-78.56
MLD	1	0.004529	0.000835	-84.76
PS	1	0.005132	0.000717	-76.56
MIA	1	0.007003	0.000973	-77.78
MMIA	1	0.00695	0.000685	-84.03
Start: AIC=-92.36				
XE~ Level + IFTI + ISD+ Age+ IT+ MLD+ PS+ MIA+ MMIA				
<none>	——	——	0.000736	-92.36
Level	1	0.004622	0.000579	-78.09
IFTI	1	0.005522	0.000932	-80.55
ISD	1	0.003179	0.000536	-81.27
Age	1	0.006673	0.000915	-77.69
IT	1	0.00728	0.000544	-73.17
MLD	1	0.007463	0.00082	-74.97
PS	1	0.003463	0.000503	-71.12
MIA	1	0.006616	0.00086	-75.24
MMIA	1	0.007125	0.00095	-80.55

Multiple nonlinear regression analysis was done on the data after eliminating the variables using R language and the results obtained are shown in Table 4. It can be seen that the overall p-value of the model (0.005243), indicates that the model can pass the significance test of the equation. The P-values (0.009-0.098) for each parameter individually are also less than 0.1, indicating that the test of significance of the regression coefficients also passes. The R-square of the regression equation changed from 0.9892 before the variables were removed to 0.9959, indicating that the regression equation became a better fit to the data. Therefore, it can be found that the number of times users browse products and product satisfaction in cross-border e-commerce Wish platform have a greater degree of influence on sales (P = 0.010, 0.009).

Table 4. Multivariate nonlinear regression analysis results after removing the variable

Independent variable	Estimate	Std.Error	t value	Pr(>|t|)
Level	0.15372	0.9978	0.999	0.079
IFTI	0.68004	0.8014	0.707	0.039
ISD	1.29069	0.7585	1.392	0.098
Age	0.76267	0.6502	1.665	0.023
IT	1.21503	0.8384	1.332	0.064
MLD	1.74724	0.9579	1.515	0.010
PS	1.52967	0.8545	1.809	0.009
MIA	1.39484	0.9203	0.465	0.032
MMIA	1.14603	0.835	1.109	0.095
Multiple R-squared: 0.9996, Adjusted R-squared: 0.9959				
F-statistic: 185.94, p-value: 0.005243				

4.3. Analysis of the effect of sales strategy optimization and adjustment

From the results of the above multivariate non-linear regression analysis, it can be seen that the main influencing factors of the sales in the cross-border e-commerce Wish platform of Company A are the number of times users browse the commodities and the satisfaction of the commodities, which indicates that Company A can start from the improvement of the degree of promotion of the commodities and the commodity's quality control rate, adjust and optimize the sales strategy to achieve the precise marketing program. After the implementation of precision marketing, the sales share of stores in the cross-border e-commerce Wish platform of Company A has also found certain changes, and the analysis results of the sales share of different types of users in Company A before and after the optimization and adjustment of the precision sales strategy are shown in Table 5. In terms of the overall situation, the share of customers in the United States (41.72%), Europe (27.79%) and Canada (14.11%) regions after the implementation is still in the top three, but the proportion has further increased, while the share of customers in Asia and other regions has further decreased. This may be related to the promotional channels of the goods, and the promotion on platforms such as TikTok further gathered customers from these developed regions, which led to the increase in the sales share. On the other hand, in terms of user types, the sales share of the three types of high-value users, energetic users and growth users in Europe and the United States region all increased compared with that before the implementation of the optimized sales strategy, which indicates that the adjustment of the sales strategy based on the results of the multivariate nonlinear regression analysis has achieved a good implementation effect.

Table 5. Different types of user sales analysis

Country	High value quality user		Dynamic quality user		Growth user		Total	
	Before	After	Before	After	Before	After	Before	After
The United States	9.56%	10.52%	4.26%	5.36%	25.64%	25.84%	39.46%	41.72%
Europe	5.69%	5.94%	5.32%	5.94%	15.48%	15.91%	26.49%	27.79%
Canada	3.69%	2.65%	1.25%	1.59%	9.65%	9.87%	14.59%	14.11%
Australia	2.09%	2.54%	0.95%	1.21%	5.24%	5.97%	8.28%	9.72%
Other	1.80%	0.95%	0.86%	0.99%	3.59%	2.31%	6.25%	4.25%
Asia	2.36%	0.85%	0.94%	0.94%	1.63%	0.62%	4.93%	2.41%

5. Conclusion

This paper cleans and integrates user data collected from cross-border e-commerce Wish platform, and constructs multivariate nonlinear regression model with product sales as the target variable, and user gender, product satisfaction and other variables as the independent variables.Company A mainly focuses on retailing on cross-border e-commerce Wish platform, with an average of 14 orders per day in the store, and an average of 313 million dollars in monthly sales. The average monthly sales is only 31.3 million dollars. Therefore, the data of Company A in the cross-border e-commerce Wish platform as the research object, analyze its user behavior data on the degree of influence on the sales of goods, and analyze the effect of the implementation of sales strategy adjustment, the results show that:

1) There is correlation between the variables selected for the study, and the correlation coefficient is between 0.212 and 0.683, indicating that the selected variables support multiple nonlinear regression analysis. The significance value of the multivariate nonlinear regression model is 0.005243, indicating that the user data in the cross-border e-commerce Wish platform has a significant impact on merchandising. In addition, the number of user views of goods and product satisfaction in the cross-border e-commerce Wish platform have a greater degree of influence on sales (P = 0.010, 0.009).

2) After adjusting the marketing strategy according to the results of regression analysis, Company A has the highest percentage of customers in the United States (41.72%), Europe (27.79%) and Canada (14.11%) regions, and the proportion has increased compared to the pre-sales strategy adjustment.

The results of the analysis of the impact of various user behaviors on product sales in the cross-border e-commerce Wish platform in this study can help cross-border e-commerce enterprises to more accurately carry out the development of market positioning and sales promotion strategies, etc., and increase the amount of merchandise marketing and brand awareness of cross-border e-commerce enterprises.

Funding

[Fund project] One of the achievements of the second batch of teaching reform projects in the 14th Five-year Plan in Zhejiang Province: Research on the innovation of e-commerce teaching system under the background of industry and education integration (project number: jg20240332);

One of the achievements of the educational scientific planning project in Zhejiang Province Exploration of enterprise customized closed-loop teaching mode under the new double-high perspective (project number: GH2025136).

References

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

[2] Xu, Y., He, D., & Fan, M. (2024). Antecedent research on cross-border E-commerce consumer purchase decision-making: The moderating role of platform-recommended advertisement characteristics. Heliyon, 10(18).

[3] Ma, S., Chai, Y., & Jia, F. (2022). Mitigating transaction risk for cross-border e-commerce firms: A multiagent-based simulation study. International Business Review, 31(4), 101965.

[4] Chen, W. H., Lin, Y. C., Bag, A., & Chen, C. L. (2023). Influence factors of small and medium-sized enterprises and micro-enterprises in the cross-border e-commerce platforms. Journal of Theoretical and Applied Electronic Commerce Research, 18(1), 416-440.

[5] Yan, Z., Lu, X., Chen, Y., & Wang, K. (2023). Institutional distance, internationalization speed and cross-border e-commerce platform utilization. Management decision, 61(1), 176-200.

[6] Fan, Z., Wang, Y., & Ying, Z. (2023). Empowerment of cross-border e-commerce platforms for small and medium-sized enterprises: evidence from China. Journal of Business-to-Business Marketing, 30(1), 33-44.

[7] Song, Y. (2021). Application of big data analysis technology in cross-border e-commerce. Journal of Electronic Research and Application, 5(4), 1-3.

[8] He, Y., & WANG, J. (2019). A panel analysis on the cross border e-commerce trade: Evidence from ASEAN countries. The Journal of Asian Finance, Economics and Business, 6(2), 95-104.

[9] Purnomo, Y. J. (2023). Digital marketing strategy to increase sales conversion on e-commerce platforms. Journal of Contemporary Administration and Management (ADMAN), 1(2), 54-62.

[10] Dai, H., Xiao, Q., Yan, N., Xu, X., & Tong, T. (2022). What influences online sales across different types of e-commerce platforms. International Journal of Electronic Commerce, 26(3), 311-330.

[11] Reddy, V. M., & Nalla, L. N. (2021). Harnessing Big Data for Personalization in E-commerce Marketing Strategies. Revista Espanola de Documentacion Cientifica, 15(4), 108-125.

[12] Rosário, A., & Raimundo, R. (2021). Consumer marketing strategy and E-commerce in the last decade: a literature review. Journal of theoretical and applied electronic commerce research, 16(7), 3003-3024.

[13] Potwora, M., Zakryzhevska, I., Mostova, A., Kyrkovskyi, V., & Saienko, V. (2023). Marketing strategies in e-commerce: personalised content, recommendations, and increased customer trust. Financial and Credit Activity: Problems of Theory and Practice, 5(52), 562-573.

[14] Pan, Y. (2024). The role of recommendation algorithms in driving the development of e-commerce platforms. Applied and Computational Engineering, 92, 206-211.

[15] Meizhi, T., & Zhongzheng, W. (2024, April). Marketing Strategy of AI E-commerce Platform Based on User Profile. In 2024 IEEE 4th International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB) (pp. 527-529). IEEE.

[16] Yang, J. (2023, October). Design Mechanism of Precision Marketing Planning Platform Based on User Data Mining. In 2023 IEEE 5th Eurasia Conference on IOT, Communication and Engineering (ECICE) (pp. 445-447). IEEE.

[17] Zhang, D., & Huang, M. (2022). A Precision Marketing Strategy of e - Commerce Platform Based on

Consumer Behavior Analysis in the Era of Big Data. Mathematical Problems in Engineering, 2022(1), 8580561.

[18] Lai, X. (2022, April). Precision Marketing Strategy Design of E-commerce Platform Based on Data Mining in the Era of Big Data. In International Conference on Multi-modal Information Analytics (pp. 1011-1017). Cham: Springer International Publishing.

[19] Sun, Y. (2021). Research on Precision Marketing of Big Data in Small-and Medium-Sized E-commerce Enterprises. In Modern Industrial IoT, Big Data and Supply Chain: Proceedings of the IIoTBDSC 2020 (pp. 121-127). Springer Singapore.

[20] Xiang Cheng,Miaomiao Kuang & Hongyu Yang. (2024). Missing Data Imputation Based on Causal Inference to Enhance Advanced Persistent Threat Attack Prediction. Symmetry(11),1551-1551.

[21] Wenyan Xu,Yucong Yan & Xiaonan Li. (2024). Sequential rough set: a conservative extension of Pawlak's classical rough set. Artificial Intelligence Review(1),9-9.

[22] Ting Lan,Rong Hu,Qi Tang,Minxia Han,Shuqin Wu & Gang Liu. (2025). A multivariate nonlinear regression prediction model for the performance of cooling tower assisted ground source heat pump system. Energy Conversion and Management119333-119333.