Big Data Analysis of User Behavior on E-commerce Platforms

Authors: Zhao Sanglin, Tong Jia'ang, Deng Hao, wandering star, Yuan Bingkun, Xinyi Huang, Zhao Sanglin

Date: 2024-08-18T00:00:00+00:00

Abstract

Statistical analysis and modeling were conducted based on user transaction behavior data from e-commerce platforms, mainly including the following aspects: 1. Statistical Analysis and Visualization of User Consumption Behavior: Statistical analysis was performed on user transaction behavior, including transaction amount, transaction frequency, transaction time, and other aspects. Through data visualization tools, charts and visualization dashboards were created to display trends, distributions, and key metrics of user consumption behavior, such as total transaction amount and average transaction frequency. These visualization results helped this study gain in-depth understanding of user consumption behavior patterns and provided a basis for further analysis and decision-making. 2. Feature Engineering and Modeling: Feature engineering was conducted based on user transaction behavior data to extract effective features. These features include user transaction amount, transaction frequency, recency of last transaction, user preferences for product categories purchased, etc. These features were utilized to construct user, merchant, and coupon features, providing a foundation for subsequent analysis and modeling. 3. User Profiling Analysis: Based on information such as user attributes, preferences, and behaviors, a user profiling model was constructed to describe users through labeling. Through user profiling analysis, we can gain in-depth understanding of user characteristics and behavior habits, such as high-value users, frequent purchasers, new users, etc. These user profile labels can provide important references for precision marketing and personalized user recommendations. 4. Merchant Voucher Distribution Model: A model was constructed and trained based on user transaction behavior data to predict whether to distribute merchant vouchers to users. Through feature selection, model training, and evaluation, a predictive model was established that can predict whether users meet the conditions for voucher distribution based on their transaction behavior. Finally, the model's performance was evaluated, including metrics such as accuracy and recall rate, to

ensure the model's effectiveness and reliability. 5. Coupon Placement Strategy Design: A coupon placement strategy was designed based on the above analysis and modeling results. This strategy considers factors such as user consumption behavior, user profiles, and the merchant voucher distribution model, aiming to achieve precise placement and improve coupon utilization rates. The strategy may include aspects such as target user positioning, discount level setting, and placement channel selection, to maximize coupon effectiveness and return.

Full Text

Big Data Analysis of User Behavior on E-Commerce Platforms

Authors: Zhao Sanglin[1], Tong Jia'ang[2], Deng Hao[1], You Xing[3], Yuan Bingkun[4], Huang Xinyi[1]
Affiliations:
[1] Hunan University of Finance and Economics, School of Economics, 410205
[2] Central South University, School of Energy Science and Engineering, 410083

Abstract

This study conducts statistical analysis and modeling based on user transaction behavior data from e-commerce platforms, encompassing five key aspects: (1) Statistical analysis and visualization of user consumption behavior, including transaction amount, frequency, and timing; (2) Feature engineering and modeling to extract effective features such as transaction amounts, frequency, recency, and product category preferences; (3) User profiling analysis to categorize users into segments like high-value, frequent buyers, and new users; (4) Merchant voucher issuance models to predict voucher eligibility based on transaction behaviors; and (5) Coupon delivery strategy design incorporating consumption patterns, user profiles, and predictive models to achieve precision targeting and maximize coupon utilization. These analyses provide actionable insights for precision marketing and personalized recommendations.

Keywords: visualization, feature selection, user portrait, normal probability graph

---

1. Statistical Analysis and Visualization of User Consumption Behavior

1.1 Statistical Analysis

We performed descriptive statistical analysis on user transaction behaviors using SPSS 26. The analysis examined transaction amounts, frequency, timing, and other key metrics. The SPSS syntax for data import and descriptive statistics is as follows:

```
GET DATA /TYPE=XLSX /FILE='C:\Users\HW\Desktop\data.xlsx' /SHEET=name 'Sheet1'
/CELLRANGE=FULL /READNAMES=ON /DATATYPEMIN PERCENTAGE=95.0 /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME 数据集1 WINDOW=FRONT.
DESCRIPTIVES VARIABLES=使用状态 邮费 购买数量 实付金额 商家ID 付款日
/STATISTICS=MEAN STDDEV VARIANCE RANGE MIN MAX.
```

The SPSS output provided descriptive statistics including mean, standard deviation, variance, range, minimum, and maximum values for key variables such as usage status, postage, purchase quantity, actual payment amount, merchant ID, and payment date.

1.2 Data Interpretation

The dataset contains the following key variables:

- **Order ID:** Unique identifier for each transaction
- **Merchant ID:** Unique identifier for each merchant
- **User ID:** Unique identifier for each user
- **Time Period:** January 2022 to June 2022
- **Usage Status:** 1 = used merchant coupon; 0 = collected but not used; None = not collected
- **Actual Payment:** Amount paid by user (in RMB)
- **User Province:** Province where user is located
- **User City:** City where user is located
- **Purchase Quantity:** Number of items purchased

1.2.1 Basic Data Visualization

We utilized STATA 17.0 to construct histograms for key variables, providing visual representations of their distributions:

```
import excel "C:\Users\HW\Desktop\data.xlsx", sheet("Sheet1") firstrow
(10 vars, 20,183 obs)
histogram 实付金额 (bin=43, start=-300, width=30.883721)
histogram 邮费 (bin=43, start=0, width=.41860465)
histogram 购买数量 (bin=43, start=1, width=1.8837209)
histogram 使用状态 (bin=42, start=0, width=.02380952)
```

These histograms visualize the distribution patterns of actual payment amounts, postage fees, purchase quantities, and coupon usage status across the 20,183 observations.

1.3.1 Provincial User Characteristics

Data visualization reveals that economically developed provinces—including Shanghai, Guangdong, Jiangsu, and Beijing—have significantly higher user concentrations on the e-commerce platform. This geographic distribution has several strategic implications:

Regional Positioning: E-commerce platforms can implement region-specific strategies based on provincial user distribution. By understanding consumption habits, purchasing power, and market demands across different regions, platforms can tailor sales strategies and product promotion plans accordingly. Regional positioning also enables optimization of logistics and distribution networks to provide faster, more efficient services.

Supply Chain Management: Provincial user data provides critical insights for supply chain optimization. Platforms can strategically plan inventory allocation and supplier relationships based on regional demand patterns, ensuring adequate stock availability and timely delivery. Understanding regional preferences also facilitates more effective inventory management and product deployment.

Marketing Strategies: Platforms can develop personalized marketing campaigns tailored to provincial characteristics. This includes launching region-specific promotions, customized product recommendations, and targeted advertising campaigns. By understanding local consumption patterns, platforms can more precisely target users, improving advertising effectiveness and return on investment.

Merchant Expansion: Geographic distribution data guides strategic merchant acquisition and partnership development. Platforms can proactively attract influential merchants and brands in regions with high user concentration, thereby expanding the availability of quality products and services that meet local demands.

1.3.2 City-Level User Characteristics

Analysis shows that users are predominantly concentrated in provincial capital cities within their respective provinces. This concentration pattern offers several strategic advantages:

Urban Positioning: Understanding the concentration in capital cities enables more precise urban-level positioning. By analyzing consumption patterns, preferences, and purchasing power in these key cities, platforms can develop city-specific sales strategies and marketing plans that align with urban user needs.

Local Partnerships: Capital cities typically exhibit higher economic development, consumption capacity, and demand. E-commerce platforms can establish partnerships with local merchants and brands in these cities to offer diverse, high-quality products and services. Collaborating with influential local businesses enhances platform credibility and drives business growth.

Logistics Optimization: The concentration in capital cities provides clear guidance for logistics network optimization. Platforms can strategically establish distribution centers and develop efficient delivery systems to provide rapid fulfillment services. City-specific logistics strategies—such as same-day delivery or evening delivery—can be implemented to enhance user experience.

Regional Promotion: Different capital cities possess unique cultural and regional characteristics that platforms can leverage for targeted promotions. By collaborating with local industries and cultural products, platforms can offer distinctive merchandise that attracts user interest and builds competitive differentiation.

---

2. User Transaction Behavior Characteristics

2.1 Normal Probability Plot

We generated a normal probability plot to examine the distribution of purchase dates:

```
pnorm 付款日期, grid
```

This plot visualizes whether the payment date variable follows a normal distribution, which is essential for selecting appropriate statistical methods for subsequent analysis.

---

3. Customer Value Analysis

We randomly sampled 100 observations to construct a multiple regression model analyzing the relationship between purchase quantity and postage fees.

3.1 Correlation Analysis

We first computed the correlation coefficient matrix for key variables and visualized it using a heatmap. The analysis examined correlations between actual payment amount, postage, purchase quantity, and coupon usage status.

The correlation matrix yielded the following results:

Note: **,** *,* *and* represent significance levels at 1%, 5%, and 10% respectively.

The correlation matrix reveals that actual payment amount—the decisive indicator of user value—shows relatively strong correlations with postage and coupon usage status. Usage status indicates whether users have collected merchant coupons.

These correlations provide actionable insights for merchant decision-making:

Postage Strategy Optimization: The positive correlation between actual payment and postage suggests merchants should optimize shipping policies to enhance purchase intentions. For instance, setting free shipping thresholds based on order values can incentivize users to increase purchase quantities. Negotiating favorable logistics rates can also reduce user shipping costs.

Coupon Usage Promotion: The correlation between usage status and actual payment indicates that coupon utilization positively impacts spending. Merchants should implement strategies to encourage coupon redemption, such as personalized coupon distribution, limited-time offers, and targeted reminders to create urgency.

Differentiated Discount Strategies: Given the correlation between usage status and payment amount, merchants can develop tiered strategies based on user segments. For users who collected but haven't used coupons, personalized push notifications with attractive offers can stimulate conversions. For low-spending coupon users, adjusting discount levels or providing additional benefits can increase their actual payment amounts.

Data-Driven Personalization: Leveraging data analytics and recommendation algorithms, merchants can deliver personalized offers based on the correlation between usage status and payment amount. Understanding individual consumption patterns enables precise matching of products and promotions to user needs, thereby increasing both purchase intention and actual spending.

3.1.1 Correlation Analysis Code

```
import pandas
from spsspro.algorithm import descriptive_{analysis}

data = pandas.DataFrame({
    "A": [1, 2, 3],
    "B": [2, 3, 4]
})

# Correlation analysis
result = descriptive_{analysis}.correlation_{analysis}(data)
print(result)
```

The heatmap visualization uses color intensity to represent correlation strength, providing an intuitive understanding of relationships between variables.

3.1.2 Linear Regression Analysis

The linear regression analysis results are as follows:

Regression Results (n=100):

Variable	Unstandardized Coefficient	Standardized Coefficient	VIF
Usage Status	1.157	0.792	0.164
Actual Payment	-0.003	0.002	-0.156

Model Fit: F = 1.287, P = 0.283
Dependent Variable: Purchase Quantity

Model Diagnostics: 1. The F-test evaluates whether the overall regression model is significant. With P = 0.295, the model does not achieve statistical significance at conventional levels, suggesting we cannot reject the null hypothesis that all regression coefficients are zero.

2. All VIF values are below 10, indicating no concerning multicollinearity issues. The model is well-specified in terms of variable independence.

3. The regression equation derived is:
   y = 1.656 - 0.099 × Postage

Linear Regression Code:

```
import numpy
import pandas
from spsspro.algorithm import statistical_{{{model}}_{{analysis}}}

# Generate sample data
data_{x1} = pandas.DataFrame({
    "A": numpy.random.random(size=100),
    "B": numpy.random.random(size=100)
})
data_{x2} = pandas.DataFrame({"C": numpy.random.choice(["1", "2", "3"], size=100)})
data_y = pandas.Series(data=numpy.random.choice([1, 2], size=100), name="Y")

# Linear regression
result = statistical_{{{model}}_{{analysis}}}.linear_{regression}(
    data_y=data_y, data_{x1}=data_{x1}, data_{x2}=data_{x2}
)
print(result)
```

---

4. Coupon Delivery Strategy

Based on descriptive statistics showing an average actual payment of 117 RMB, we propose a coupon strategy anchored to this benchmark:

Fixed-Amount Coupons: Offer fixed discounts (e.g., 50 RMB or 100 RMB) to incentivize purchases. A 50 RMB coupon could target users at the 25th percentile, while a 100 RMB coupon targets the top 10% of spenders.

Percentage Discount Coupons: Provide percentage-based discounts (e.g., 10% or 20% off) that scale with purchase amount, encouraging larger orders.

Threshold-Based Coupons: Implement "spend-and-save" promotions such as "200 RMB minus 50 RMB" or "300 RMB minus 100 RMB" to motivate users

to increase their basket size.

Tiered Coupon Structure: Create multiple discount tiers based on spending levels: modest discounts for purchases under 120 RMB, moderate discounts for 120-200 RMB, and substantial discounts for orders exceeding 200 RMB.

Personalized Coupons: Leverage purchase history and preference data to deliver customized offers. High-value loyal customers could receive exclusive discounts to strengthen retention and increase purchase frequency.

---

5. Research Significance

Analyzing e-commerce user big data offers several critical contributions:

Consumer Behavior Insights: E-commerce platforms accumulate vast behavioral data including browsing, search, purchase, and review activities. Analyzing this data reveals purchasing patterns, preferences, interests, and needs, illuminating decision-making processes and motivations. These insights guide product development, marketing strategies, and supply chain management.

Personalized Recommendation Optimization: User big data enables refinement of recommendation algorithms. By analyzing historical purchases, browsing behavior, and interests, platforms can deliver tailored content that enhances shopping experience and satisfaction, ultimately increasing conversion and retention rates.

Customer Segmentation and Relationship Management: Data mining and machine learning techniques facilitate user segmentation into distinct groups (high-value, potential, loyal customers). Segment-specific strategies improve customer satisfaction and loyalty through targeted marketing and relationship management.

Predictive Analytics and Demand Forecasting: Historical data analysis uncovers purchasing patterns and trends, enabling accurate demand forecasting. This supports inventory planning, procurement decisions, and supply chain optimization while minimizing operational costs.

Fraud Prevention and Security: Analyzing user data helps build anti-fraud models and security systems to identify and prevent fraudulent activities, protecting user rights and data security.

Overall, e-commerce user big data research is essential for optimizing personalization, customer relationship management, and marketing strategies. It enhances user satisfaction and loyalty, drives business growth, and strengthens market competitiveness while improving operational efficiency and risk management.

---

6. References

[1] Wang Zhibo. Research on E-commerce Platform User Data Mining Based on Hadoop[D]. North China Electric Power University, North China Electric Power University (Baoding), 2021.
[2] Ji Nanxun. Multi-source Heterogeneous Data Fusion and Representation for E-commerce Platform User-Generated Content[D]. China University of Mining and Technology, China University of Mining and Technology (Jiangsu), 2021.
[3] Xu Xiwei, Zhang Weirong, Chang Houfeng. Construction and Optimization of E-commerce User Behavior Big Data Platform[J]. Electronics, 2022(5): 93-95. DOI:10.3969/j.issn.1006-4052.2022.05.028.
[4] Shan Wenfa, Xiao Yi, Nie Xiaoyi. Research on the Relationship Between Data Mining and Improving User Stickiness on E-commerce Platforms[J]. China Business Review, 2016(18): 66-67. DOI:10.3969/j.issn.1005-5800.2016.18.035.
[5] Zhang Lei. Evolutionary Game Between Users and E-commerce Platforms' "Big Data Price Discrimination"—Based on User-Government Collaborative Supervision Mechanism[D]. University of Chinese Academy of Sciences, 2022.
[6] Li Peng. Design and Implementation of User Digital Portrait System for Publication E-commerce Platform Based on Big Data[D]. Beijing: Beijing University of Technology, 2018.
[7] Peng Geng, Zhang Lei, Han Chenyang. Evolutionary Game of "Big Data Price Discrimination" Between Users and E-commerce Platforms[J]. Science and Technology for Development, 2022, 18(3): 367-374. DOI:10.11842/chips.20211015001.
[8] Guangzhou Jiuling Blockchain Technology Co., Ltd. A Blockchain-based E-commerce Platform User Data Query Method with Information Theft Prevention: CN202010768341.9[P]. 2020-11-27.
[9] Liu Yang. Research on Statistical Analysis Methods for E-commerce Platform User Review Text Data[D]. Heilongjiang: Heilongjiang University, 2021.
[10] Jiang Lihua, Shen Jinyu, Ren Yi. Sentiment Analysis of E-commerce Platform User Review Data Based on LDA Topic Analysis Model—Taking Fresh Agricultural Product Apple as an Example[J]. Technology and Market, 2021, 28(10): 74-76. DOI:10.3969/j.issn.1006-8554.2021.10.023.
[11] Zhang Xiaoying, Zhang Qiuhong, Yang Xinwei. Big Data Analysis of E-commerce Platform User Purchase Intention[J]. Journal of Changchun University (Natural Science Edition), 2020, 30(3): 6-8, 27.
[12] Zhang Yunhao. Issues and Solutions in the Trial of E-commerce Platforms' Unreasonable Use of User Data Infringement Cases[J]. Legal System Expo, 2019(22): 194-195. DOI:10.3969/j.issn.2095-4379.2019.22.098.
[13] Gong Guoteng. A Precise Service Management Method for E-commerce Platform Users Based on Big Data[J]. Industrial Innovation Research, 2021(8): 26-28.
[14] Qi Yingping. Application of Data Mining in User Behavior Analysis of a Cross-border E-commerce Platform[D]. University of Chinese Academy of Sciences, 2020.

[15] Lu Chengcheng. Design and Implementation of E-commerce User Behavior Big Data Analysis Platform[D]. Shandong: Ocean University of China, 2019.
[16] Huang Renwei. User Linkage Analysis Method and System for E-commerce Big Data Platform: CN202211653608.5[P]. 2023-03-07.
[17] Kunming Yuanzengdai Information Technology Co., Ltd. User Online Intention Analysis Method Based on Artificial Intelligence and Big Data E-commerce Platform: CN202211637844.8[P]. 2023-03-31.
[18] Sichuan Public Information Industry Co., Ltd. An Internet E-commerce Platform User Privacy Data Encryption Method: CN202310670118.4[P]. 2023-09-05.
[19] Zhengzhou Lizhen Information Technology Co., Ltd. E-commerce Platform User Classification Method Based on Big Data: CN202211297618.X[P]. 2023-01-31.
[20] Shandong Foreign Trade Vocational College. A Big Data User Mining Method Based on Cross-border E-commerce Platform: CN202210837876.6[P]. 2022-10-14.

Note: Figure translations are in progress. See original paper for figures.