Hindawi
Advances in Multimedia
Volume 2022, Article ID 4291978, 22 pages
https://doi.org/10.1155/2022/4291978

Research Article

Visual Analysis of E-Commerce User Behavior Based on Log Mining

Tingzhong Wang, Nanjie Li, Hailong Wang, Junhong Xian, and Jiayi Guo

School of Information Technology, Luoyang Normal University, Luoyang 471934, China

Correspondence should be addressed to Tingzhong Wang; wangtingzhong@lynu.edu.cn

Received 29 January 2022; Accepted 15 April 2022; Published 5 May 2022

Academic Editor: Qiangyi Li

With the continuous development of internet economy and e-commerce, the scale of data produced by users on e-commerce platform is increasing explosively. Mining the behavior of individual users and group users from massive user behavior data and analyzing the value and law behind the data are of great significance to the development of e-commerce. Taking the user behavior log data of an e-commerce website as the data source, this paper, firstly, processes and analyzes the original dataset through the data filtering and storage module, and it uses the combination of Kafka and Flume to store the user behavior log with reasonable structure and complete fields in HDFS. Secondly, a hierarchical system of data warehouse is constructed in Hive, and each layer of log data is effectively mined and multidimensionally analyzed with the help of log mining technology. Finally, based on the big data framework and Bi tools, a data warehouse system is designed and implemented, which could store and analyze massive data and visually display the results. The system uses dimensional modeling to build a data warehouse hierarchical system to mine and analyze user behavior data through log mining algorithm deeply. The *K*-means clustering algorithm and RFM model are used to divide the user behavior characteristics in detail, and AARRR funnel model is used to analyze the logs in a modular way. Through the effective mining and multidimensional visual analysis of user behavior data, the behavior analysis of group users and individual users, as well as the analysis of commodity sales flow and sales linkage are realized, which provides support for internal decision-making and precision marketing.

1. Introduction

Under the "internet plus" policy and the new normal economic development, China's online retail sales in the first half of 2021 were 61,133 billion yuan, up 23.2% over the same period last year. With the in-depth research and rapid development of internet technology, the internet has gradually become an integral part of people's daily life. E-commerce, especially e-commerce platforms, such as jd.com, Taobao, and Pinduoduo, has attracted more people to choose online shopping because of the efficient and convenient commodity transaction methods. The access situation, transaction process, address location, and system status generated by many users on the e-commerce platform are recorded in the form of logs. Every e-commerce platform generates massive amounts of log data all the time. The timely identification of emerging trends from massive amounts of data plays an important role in business processes and decision-making. Digital operation of e-commerce enterprises could be realized by effective mining and visual analysis of log data; thus the enterprises could timely and accurately understand the real needs of customers, and make predictions about future operations in advance, so that the cost of enterprises could be minimized and profits could be maximized.

Clustering is an important technology of e-commerce log mining, which could be roughly divided into four categories: partition method, hierarchical method, density method, and grid method. Clustering can be applied to information retrieval, web page grouping, and image and market segmentation. Customer market segmentation is an

effective method in e-commerce log mining. To conduct segmentation identification, customer characteristics and behavior characteristics related to products (such as purchase behavior, consumption behavior, preference for goods, experience, and service) are used to conduct customer clustering. The *K*-means clustering algorithm is a very important technology of data mining. In practical application, it mainly divides the sets into different types according to similarity function and similarity criterion, and it minimizes the differences existing in the same types. If the differences in each category are described by the maximum cluster analysis, it can be understood as determining the number of internal vectors in the M-dimensional space, classifying all vectors into one of several clusters and minimizing the distance between all vectors and the cluster center. The RFM (last transaction, frequency, and monetary value) model is an important tool and means for the e-commerce platform to measure the current user value and customer potential value. Customers are classified according to the RFM model. User groups are divided based on the actual purchase behavior data of users and then divided into different groups for operation based on different classification information so that enterprises can more effectively obtain customers, make customers more satisfied, retain customers, become high-value customers, and avoid customer loss. In log mining, the *K*-means clustering algorithm and RFM model are used to cluster the customers and divide user behavior characteristics in detail, which can effectively improve the efficiency of data analysis and collection and extract valuable information from massive log data to build a perfect log mining system to quickly complete log mining, and providing users with the required information and improving the level of log mining is very important.

Data mining can obtain information from log data, but how to display this information quickly and effectively is another difficulty. Data visualization can present the hidden laws and features in the data in a graphical way so that people can quickly and intuitively understand the information in the data and improve their cognition and exploration ability of the data. The e-commerce log has a strong time correlation, belonging to time sequence information, which contains commodity category information with hierarchical structure and multidimensional attributes, and the user's location is geographic information. Users' purchasing behavior is not only related to commodity prices, preferential activities, and other events but also related to users' own attributes (such as interest points and purchasing power). However, traditional visualization methods fail to fully combine the characteristics of e-commerce logs, such as time sequence, region, and level, to achieve the multidimensional analysis of individual user behavior and group user behavior.

This paper focuses on the above problems and designs and implements a data warehouse visualization system that can store and analyze massive data and visually display the results. The system has four modules: data filtering and storage module, data warehouse and log mining module, data visualization module, and ad hoc query module. The data filtering and storage module processes and analyzes the original dataset, extracts the user behavior log data using Flume, sets the ETL interceptor and the log type interceptor during the extraction process, and stores the user behavior log with reasonable structure and complete field in HDFS by combining Kafka and Flume. The data warehouse adopts hierarchical architecture design, and it constructs a hierarchical system of data warehouse in Hive. With the help of log mining technology, the log data of each layer is effectively mined and multidimensionally analyzed, and the effective data is sent to the next layer. Finally, it is summarized in the ADS layer. The data visualization module, firstly, extracts the data of the ADS layer of the data warehouse into the MySQL database. Then, it designs and realizes data visualization with the help of the third-party BI tools and selects users, traffic, members, goods, sales, and other e-commerce core topics for report presentation. By deploying a Kylin Cluster, the AD Hoc query module returns multidimensional analysis results in subseconds. At the same time, a number of visual analysis methods are designed, including the multidimensional composite sequential visual analysis method, multidimensional spatiotemporal visual analysis method, and multidimensional linkage visual analysis method, based on log mining.

Figure 1 is the main interface of the system realized in this paper, and the main contributions of this paper are as follows:

(1) Selected the current mainstream big data framework and BI tools to design and implement a data warehouse system that can store and analyze massive data and display the results visually.

(2) Designed a multidimensional composite time series visual analysis method based on log mining and explored the data change trend in different time modes and different time granularities.

(3) Designed a multidimensional space-time visual analysis method based on log mining, which can effectively analyze the changes of data in different dimensions and comprehensively display user behavior data.

(4) A multidimensional linkage visual analysis method based on log mining is designed. Through data linkage, driller, dimension switch, and other analysis operations, linkage multigraphs are analyzed around the same topic.

2. Related Work

In today's e-commerce market, most market analysis uses market segmentation to simulate business policies. Different market segments provide consumers with goods and services to improve competitiveness [1], which is a huge potential business opportunity and requires relevant visualization strategies for this scenario [2–4]. In this case, existing visualization tools are only used as the static descriptions of measurement dashboards and organizational networks rather than as exploratory mining, while the market requires interactive commercial tools, and the circular segmentation technique with one color pixel per data value is considered to be a powerful expressive tool for

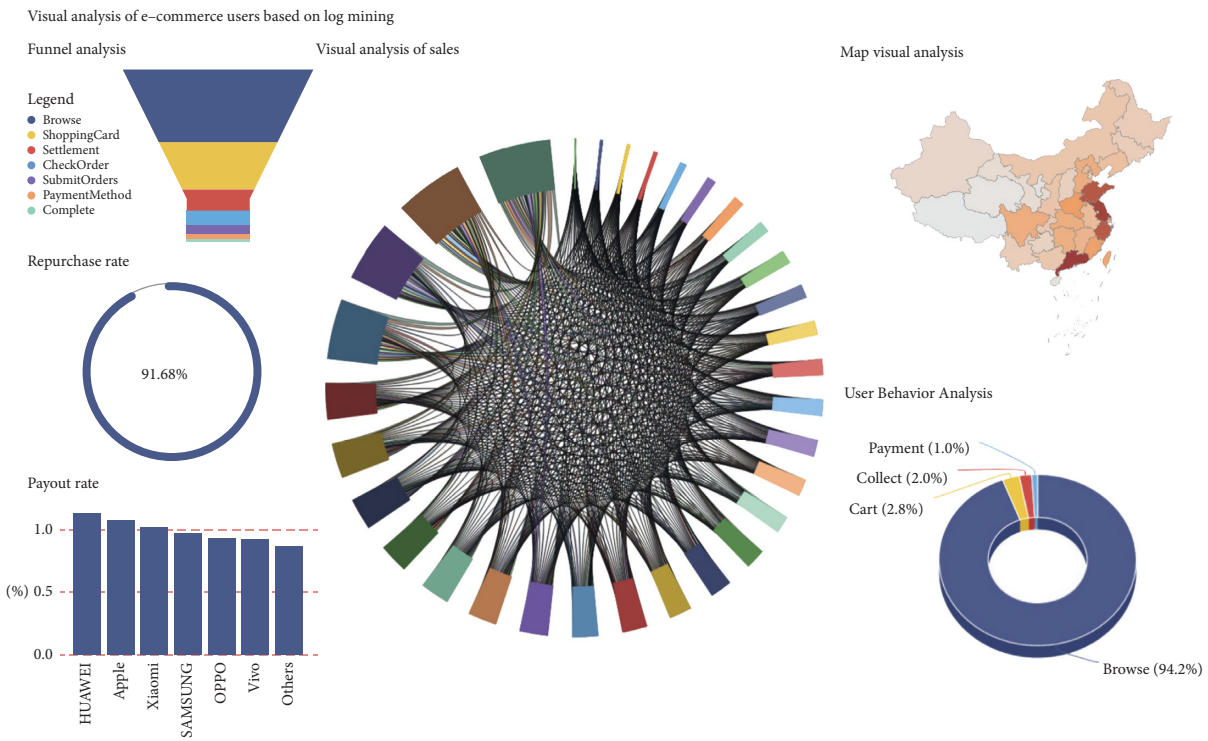


Figure 1: Overview of the proposed system.

displaying large amounts of high-dimensional data [5, 6]. To find potential customers, market segmentation is a common marketing strategy. It divides consumers with common needs into groups or subsets, and it carries out different marketing mix for different users [7–9]. Market segmentation contributes to the formulation of market diversification strategies to maximize profit margins [10, 11]. Customer segmentation improves revenue through precision marketing, especially in the way of pop-up push, SMS e-mail, and so on. Market segmentation is a customized method that captures customer groups through customized plans and programs.

In the face of the increasing amount of information, consumers buy online. Using a data clustering algorithm to segment the user market could effectively solve many problems. Clustering refers to the unsupervised process of classifying a large number of data elements into separate homogeneous clusters based on similarity. Although it is promising in many application fields, such as model classification, data mining, and decision making, decision makers need to develop some constraints to exert the effectiveness of the algorithm in the case of little prior knowledge of data properties [12]. Therefore, considering these constraints, choosing an appropriate method for effective exploration of the relationship between data elements for effective assessment is very important; usually by global optimization (on all models) or local (on the model subset), define the standard function to generate a cluster. A simple and common algorithm is K-Means algorithm [13]. The algorithm starts with an initial random partition and continuously reassembles the data into the cluster according to the similarity between the pattern and the cluster However, the main disadvantage of this algorithm is that when a large range of log data is encountered in the practical application, it tends to converge to the local minimum, which leads to the failure of effective segmentation of customer groups. To solve this problem, this paper proposes a method combining the k-means clustering algorithm and the recency frequency monetary (RFM) model, which could realize the effective division of user behavior characteristics.

The RFM models are known as behavior-based data mining techniques that extract customer data using the latest recency, frequency, and monetary value [14, 15]. In the RFM model, R represents the most recent consumption time, F represents the consumption frequency, and M represents the monetary expenditure [16, 17]. In recent years, some studies have used the concept of RFM to create customer segmentation models [18–20]. Furthermore, some researchers have tried to improve the concept of RFM by adding additional features or using data extraction techniques [16, 21, 22]. In this paper, an RFM model based on contour coefficient optimization is proposed, the K-means clustering algorithm is used to determine the optimal contour coefficient, and then RFM is used to realize the detailed division of user behavior characteristics. In the process of customer segmentation, customers are divided into different groups according to their purchasing behavior, demographics, behavioral preference, and geographical location, which is providing data support for the visual analysis of user behavior. As the number of online consumers increases, online transaction data becomes a new source of value. Some scholars use mathematical models, analytical techniques, and visualization methods to analyze consumers' purchasing behavior.

Online transaction data contains various types of attributes, such as value, time, and category. Sparklines [23] could be used to visualize multiple trends in financial data. Liu et al. [24] proposed a visualization system named SellTrend for analyzing the travel purchase requests of airlines. Chang et al. [25] proposed a visualization system for searching predefined patterns in large wired transaction datasets, and Hao et al. [26] proposed a pixel-based bar graph to analyze the correlation between various attributes of transaction logs, such as transaction time, transaction volume, and total transaction amount. By comparing multiple bar graphs, analysts can find the most valuable customers, the best time to sell, and the most searched keywords, however, the bar chart has limited dimensions and cannot display a large amount of transaction information and consumer information as the number of consumers, the number of products, and the length of time increases. Keim et al. [27] improved the pixel-based bar chart and designed the value unit bar chart, which is used to visually display the overall overview and details of the transaction log. Users can intuitively evaluate the distribution and relevance of transactions, clearly define high-value transactions and outliers, and the transaction value is immediately displayed at the transaction record level. However, the value unit bar chart does not consider the characteristics of goods and users related to the user's purchase behavior.

Chen et al. [28] designed and implemented a new visualization analysis method, namely the urban data visualization analyzer (Vaud), which supports visualization, query, and exploration of urban data. In this way, analysts can select, filter, and aggregate from multiple data sources and extract hidden information into a single data subset. Through intelligent data analysis, the method improves people's ability to identify multidimensional spatiotemporal data. Skyline query is widely used in tourism, retail, human resources, and other multistandard decision-making fields. In addition, many visualization methods for analyzing user behavior data focus on exploring the temporal patterns of individual behavior. For example, TimeSearcher [29] allows users to select a time series of interest using a rectangular query area. Plaisant [30] displays health-related events along a time line, and density-based display technology [31, 32] can display large time series datasets monitored in real time. Naeem [33, 34] explored the application of process mining in biomedical domain through real-time case study of hepatitis patients, and Qi et al. [35–38] present a trust-based collaborative filtering algorithm to perform basic rating prediction in a manner consistent with the existing CF methods, as the inherent drawbacks render preference prediction infeasible for cold-start users and have become a crucial issue to be resolved in recommendation systems. In summary, we find that the research on e-commerce user behavior log mainly focuses on the visualization of time series data, while ignoring the multidimensional attributes, hierarchical structures, and the user's own attributes of e-commerce user behavior log data. However, this information is closely related to user behavior and is important. The multidimensional data visualization exploration for e-commerce user behavior is a very meaningful work, and thus, we try to effectively analyze the behavior of e-commerce users from multiple perspectives in this paper.

3. Data Sources and Visualization Tasks

3.1. Data Sources. The data in this paper comes from the user behavior log data of Taobao Mall provided by Ali Yun Tianchi [39], and the original data includes user information, commodity information, merchant information, evaluation information, and transaction information. This paper extracted the user behavior information, commodity information, and transaction information, uploaded the data to database, and preprocessed the data through ETL, and the preprocessed data totaled 12,357,000 logs with a time span of 5 years, involving 100,000 users and 32,000 items of commodities.

User behavior information includes information, such as user ID, user nickname, mobile phone brand, item identification ID, product category ID, user behavior type, longitude and latitude, and behavior occurrence time. Among them, mobile phone brands mainly include more than a dozen brands, such as Apple, Huawei, Xiaomi, OPPO, VIVO, Samsung, etc. The latitude and longitude cover thirty-one provincial-level administrative regions in mainland China, Hong Kong, Macao, and Taiwan regions. There are seven types of user behaviors, including browsing goods, collecting goods, adding the collected goods to the shopping cart, submitting orders, selecting the payment method, completing the payment, and returning the products.

Commodity information includes commodity identification ID, commodity classification ID, commodity name, commodity brand ID, price, weight, and other information. Commodity information has hierarchical structure and multi-dimensional attributes; among them, the commodity classification information has five levels, including commodities level classification, secondary classification, commodity tertiary classification, the classification of the commodities level four categories, and category five (goods), such as household (topic)-study (scenario)-glass (material)-Chinese style (style)-Jingdezhen vase, each category can be viewed as a dimension, and each downward classification adds a corresponding dimension.

The transaction information includes information, such as serial number, order number, user ID, transaction content, transaction amount, payment type, and transaction time. The transaction log of Taobao mall is the time series data, and the time range of the sample data is from December 28, 2009, to December 28, 2014. Each log data defaults to one user transaction behavior. After preprocessing, the statistical results of user transaction behaviors are obtained. In the results, we found that the number of users with more transaction behaviors is far less than the number of users with fewer transaction behaviors, and a long tail phenomenon occurs. As the transaction behavior records of long tail users are too small, it is difficult to reflect the changing trend of users' transaction behavior. Therefore, this paper focuses on selecting 10,000 users among all users for research, using clustering algorithm to classify the

selected users, and analyzing the behavior of the users of different levels in different time patterns and different time granularities.

3.2. Visualization Tasks. By discussing with different types of electronic business, we attempt to analyze individual user behavior, group user behavior, and product sales, such as user behavior change rule, whether there is a certain time, whether the user's behavior changes are associated with the change of geographical location, whether the user has purchase preference, whether the user has one or more of the trading behaviors, how long the time interval between each transaction is, and who is the valuable value user. In addition, our tasks also include the characteristics of different user groups, the correlation between user behavior affected by promotional activities on the platform and the purchase preferences of different user groups (which goods sell best), and the relationship between product sales and time.

4. Overview of the Proposed System

For the massive e-commerce log data, the data mining algorithm is used to extract effective information, and the multidimensional data visualization method is used to display the behavior of users (individual users and group users). The system has four modules as shown in Figure 2, which are the data filtering and storage module, data warehouse and log mining module, data visualization module, and ad hoc query module. The data filtering and storage module processes and analyzes the original dataset, extracts the user behavior log data using Flume, sets the ETL interceptor and the log type interceptor during the extraction process, and stores the user behavior log with reasonable structure and complete field in HDFS by combining Kafka and Flume. The data warehouse and log mining module is used to construct the hierarchical architecture and log mining of data warehouse, and build the hierarchical system of data warehouse in Hive. With the help of log mining technology, the log data of each layer is effectively mined and multidimensionally analyzed, and the effective data is sent to the next layer, and finally, it is summarized in the ADS layer. The data visualization module extracts data from the ADS layer of the data warehouse into the MySQL database and then designs and realizes data visualization with the help of third-party BI tools. The core indicators of e-commerce, such as users, traffic, members, goods, and sales, are selected for display. By deploying a Kylin Cluster, the AD Hoc query module returns multi-dimensional analysis results in subseconds.

5. Data Mining and Data Analysis

In this paper, the *K*-Means clustering algorithm based on center point optimization and the RFM model based on silhouette coefficient optimization are used to divide the user behavior characteristics in detail, and the AARRR funnel model is used to modularize the logs and mine valuable information in the logs.

5.1. K-Means Clustering Algorithm Based on Central Point Optimization. According to the standard *K*-means algorithm, for a given sample set, the first choice is to determine the number of cluster *K* so that the samples in the cluster are distributed together as closely as possible, and the distance between the clusters is as large as possible. The algorithm attempts to divide the cluster data into n groups of independent data samples so that the variances between the groups of clusters are equal, mathematically described as minimizing the inertia or the sum of squares within the cluster. As an unsupervised clustering algorithm, *K*-means has relatively simple implementation and good clustering effect, and it is widely used in customer behavior clustering, news comments, etc. Table 1 shows the comparison between *K*-means and other clustering algorithms.

In this paper, when the standard *K*-means clustering algorithm is used for the clustering of e-commerce user behavior log data, we found that the selection of the initial center point will have a great impact on iteration efficiency and iteration effect. If the initial center point is very close, it is necessary to complete the clustering process through multiple iterations. As shown in Figure 3, if the distribution of the sample points is symmetric and the selection of initial center points is symmetric, the result of clustering is likely that half of the upper and lower distributions belong to the same cluster, which is obviously unreasonable.

Aiming at this problem, this paper adopts the *K*-Means algorithm based on center point optimization, and it optimizes the random selection of *K*-means initialization center point. For a given sample set $D = \{x_1 x_2 x_{3\ldots} x_m\}$, the process can be described as follows:

(1) Randomly select a point from the input sample set as the first cluster center point μ_1.

(2) For each point x_i in the sample set, calculate its distance from the nearest cluster center among the selected cluster centers.

(3) Select a new data point as the new clustering center. The principle of selection is as follows: the point with larger $D(x)$ has a higher probability of being selected as the clustering center.

(4) Repeat steps 2 and 3 until the k cluster center points are selected, and use these k center points as the initialization center points to run the standard *K*-means algorithm.

In this paper, four basic actions of users are selected for cluster analysis to observe the characteristics of user behavior. Because of the excessively high dimension of clustering, the spatial distance will become sparse. This paper, firstly, conducts a preliminary cluster analysis on Click, Collect, Cart, and Payment actions of 10,000 user data, and it uses the sum of the distance squares in the cluster group to find the optimal cluster number K value of 4. It divides the data into six class clusters, and the clustering results are shown in Table 2.

Click field is the total number of click operations of users. According to the clustering algorithm, 10,000 users are divided into five clusters, including 162, 842, 2657, 4574,

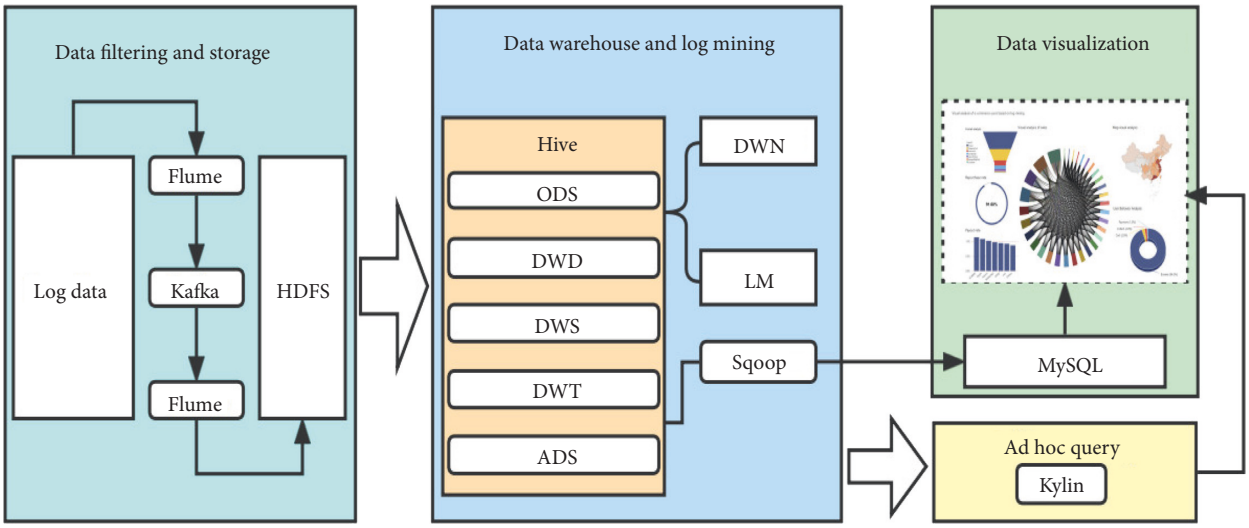


Figure 2: Overview of the system architecture.

Table 1: Comparison between *K*-means and other clustering algorithms.

Algorithm	Feature	Disadvantage	Description
CLIQUE	The speed is independent of the number of data objects and only depends on the number of cells in each dimension in the data space.	Parameter sensitive, unable to deal with irregularly distributed data, dimension disaster, etc.	The CLIQUE algorithm cannot meet the requirements of e-commerce platform by exchanging efficiency for accuracy.
FCM	The clustering effect will be very good for the data satisfying the normal distribution, and the algorithm is sensitive to outliers.	FCM cannot be guaranteed to converge to an optimal solution, and the performance of the algorithm depends on the initial clustering center.	The K-means algorithm has fast clustering speed and good clustering effect.
DBSCAN	Irregularly shaped clusters can be resolved, and it handles noisy data well.	For clusters with different densities, the DBSCAN algorithm may not work very well.	The parameter selection of DBSCAN algorithm requires manual intervention, which is slower than that of K-means.

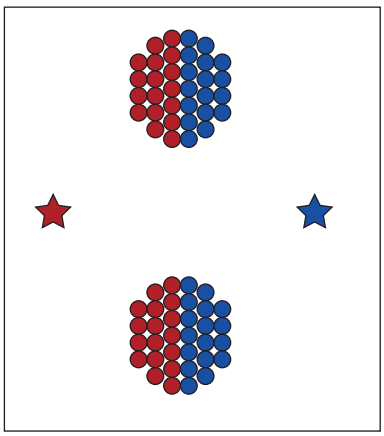


Figure 3: Symmetrical distribution of sample points.

and 1765, respectively. According to the clustering results, users can be subdivided into user groups with different active frequencies. Collect character refers to the total number of collection operations of users, and the cluster analysis of 10,000 users is conducted to obtain 5 class clusters, including 15, 460, 55, 1428, and 8042 users, respectively. By analogy with the clustering result of payment field, it is found that a large number of collection users do not generate payment. We could subdivide these users into different levels of collection hobby user groups. The cart field is the total number of shopping cart operations of users, and the cluster analysis of 10,000 users obtains 5 class clusters, including 389, 445, 47, 6116, and 1,893 people, respectively. By comparing the number of goods added to the shopping cart with the payment time of goods, we can subdivide users into user groups with different consumption modes. Commodity field refers to the total number of commodities purchased by users, and the cluster analysis of 10,000 users results in 5 clusters, including 356, 122, 755, 2430, and 6337 users, respectively. The users could be subdivided into user groups with preference for different commodities. The Churn_rate field is the browsing miss rate of users, and the clustering analysis of 10,000 users was conducted to obtain 5 class clusters. By comparing the miss rate with the user payment operation, we find that cluster 2 had a high jump rate, and casual browsing is the biggest characteristic of this type of users. Also, the clustering of the user groups with a higher jump rate of 0.3 has low shopping tendency and sincerity. We could divide users into different value groups.

Table 2: Cluster analysis results using the optimized K-means algorithm.

Type		Cluster					
		1	2	3	4	5	6
Click	User_number	10,000	162	842	2657	4574	1765
	Clicks	1155	8205	3731	1637	597	264
Collect	User_number	10,000	15	460	55	1428	8042
	Favorites	25	1287	223	60	27	11
Cart	User_number	8890	389	445	47	6116	1893
	Goods	25	218	27	25	24	11
	Duration	28	38	145	365	56	9
Payment	User_number	10,000	55	1428	15	460	8042
	Payments	12	150	36	28	20	7
Commodity	User_number	10,000	356	122	755	2430	6337
	Types_of_goods	12	55	35	20	14	7
Churn_rate	User_number	10,000	101	1236	2638	2048	3977
	Payments	12	2	11	15	12	8
	Churn_rate	0.1184	0.7328	0.3124	0.1714	0.047	0.032

5.2. RFM Model Based on Contour Coefficient Optimization. The RFM model is a statistical method for customer value segmentation, including three variables: R (recency), F (frequency), and M (monetary). The RFM model is a three-dimensional model that divides the value of e-commerce users using the three-dimensional indicators of R (last consumption time), F (consumption frequency), and M (consumption amount). Before the actual calculation, it is necessary to make two points for the three-dimensional index. For the standard RFM model, the setting method of the segmentation point is relatively simple. Usually, the size of the segmentation point is set according to the business experience of the analyst, which leads to low accuracy of the segmentation point and inaccurate user segmentation. To optimize this problem, this paper designed an RFM model based on contour coefficient. Firstly, all users were put into multiple groups of clusters. Then, different segmentation points were set. The clustering analysis of multiple groups of clusters was carried out, and the clustering effect was evaluated according to CH (Calinski Harabaz), contour coefficient, and anamorphic score. The higher the CH and anaesthesia scores were, the better the anaesthesia coefficient was, and the better with the closer the contour coefficient was to 1. Figure 4 shows the results of cluster analysis for users who purchased a certain category. In summary, it was found that cluster 3 had the best effect and the most accurate user segmentation. Hence, this method helped analysts improve the accuracy of setting segmentation points.

Through clustering analysis, our split point to R is x. The split point of F is y, and the split point of M is z. Then, we give the formula for calculating the total score of RFM, shown as (1). By multiplying with 100-10-1 and adding them, let R, F, and M be expressed as the hundreds, tens, and ones of a three-digit number, respectively.

$$\mathrm{RFM}_{\mathrm{score}} = R_S \times 100 + F_S \times 10 + M_S, \tag{1}$$

$$\begin{aligned} R_S &= \mathrm{IF}\left(R_i > x, 0, 1\right), \\ F_S &= \mathrm{IF}\left(F_i > y, 0, 1\right), \\ M_S &= \mathrm{IF}\left(M_i > z, 0, 1\right), \end{aligned} \tag{2}$$

R_S (Recency Score) indicates that the user has consumed more than X times within a year, and the user is considered to be an active user of the mall. F_S (Frequency Score) means that the consumption frequency of the user exceeds Y times within one year, indicating that the user has a high consumption frequency in the mall. M_S (Monetary Score) means that the total consumption of the user exceeds Z within one year, indicating that the user has generated high benefits for the platform. The three bits of this three-digit number represent the coordinates of the three dimensions. Figure 5 shows user classification based on the improved RMF model, and Figure 6 shows group value classification.

5.3. AARRR Funnel Model Based on User Behavior Change. The AARRR model is a new online sales model proposed for the changes in the user behavior of e-commerce platforms. Compared with other models, the AARRR model covers the user from the seeding time to the growth and harvest time of the full life cycle, and it could obtain the omnidirectional user behavior change to build effective closed-loop digital operation. On the other hand, the AARRR model emphasizes the driving role of users in the links of "retention" and "recommendation," facilitating the behavioral transformation of users from passively receiving information to actively disseminating information. This paper designs and implements the AARRR funnel model based on user the behavior changes to conduct modular analysis of logs. User groups are divided into acquisition, activation, retention, revenue, and self-propagation, as shown in Figure 7. This paper carries out the following visual analysis for each stage to obtain the number of newly added users, daily number of unique visitors, and daily clicks on e-commerce platform within the user stage, user activity in a period (such as daily, weekly, and monthly), and the time period of user preference. Improve user retention during the retention phase (day 2, 7, 14, and 30), increase the daily most popular item types, TOP10 items, and the user value during the revenue stage, and increase the user behavior conversion rate during the self-propagation stage (browse > intended purchase- > buy).

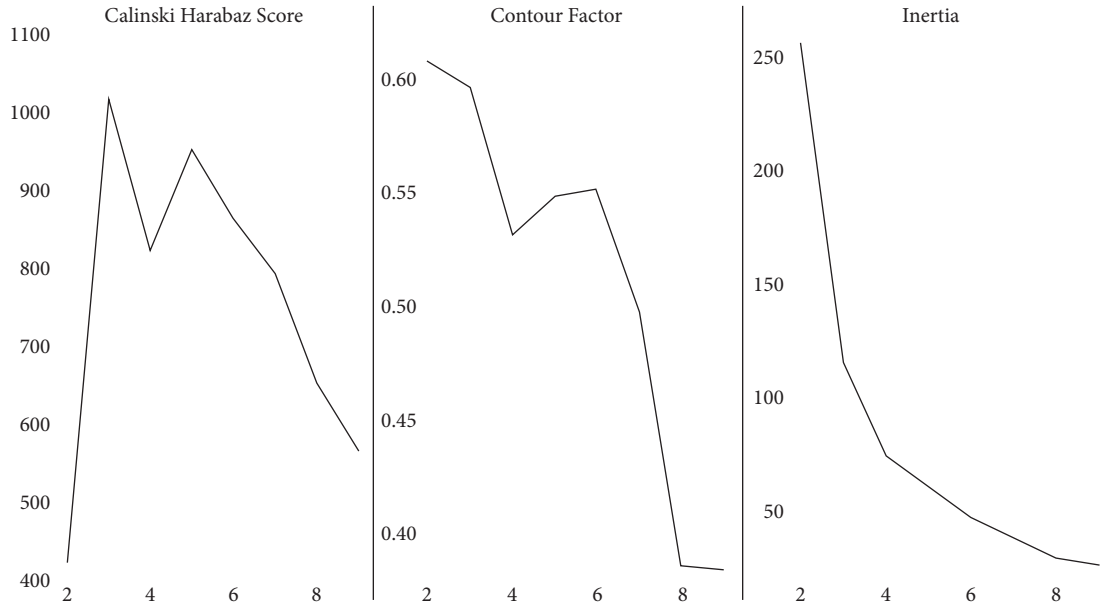


Figure 4: Clustering segmentation points.

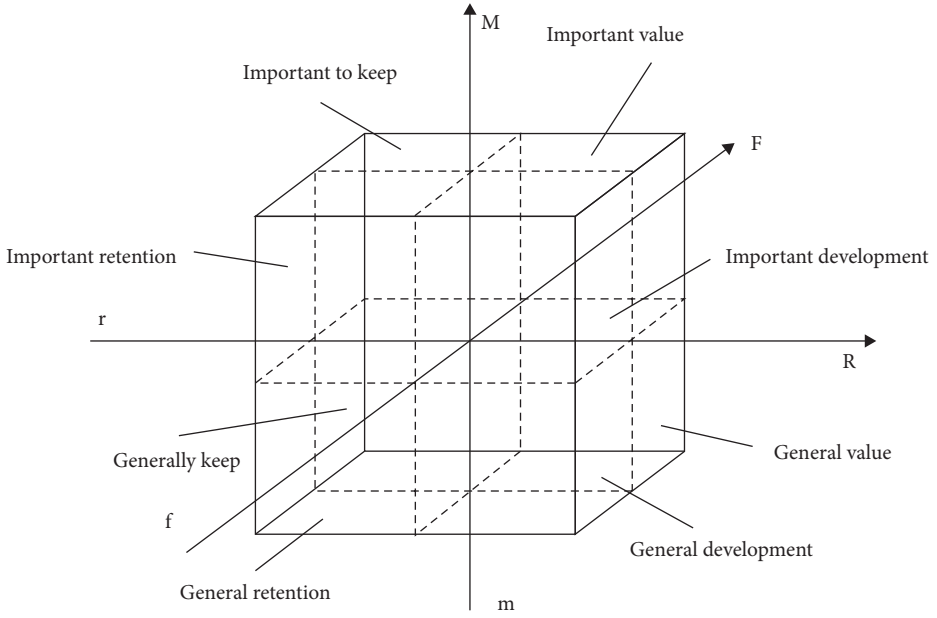


Figure 5: Classification of improved RFM 3D model.

6. E-Commerce User Behavior Visualization Method

Based on the design principle of visualization task and visualization method, this paper designs and implements the visualization method of e-commerce user behavior, including multidimensional composite sequential visual analysis method, multidimensional space-time visual analysis method, and multidimensional linkage visual analysis method based on log mining.

6.1. Multidimensional Composite Time Series Visualization Method Based on Log Mining. This paper designs and implements a multidimensional composite sequential visualization method based on log mining, which is mainly used to analyze the changing trends of individual users, group users, and commodity sales. This method solves the user behavior changes in different time granularities and different time modes and helps analysts grasp the change trend in time. For individual users, based on the composite sequential visualization method, the operation times of users are described according to different time spans (day, week, month), and the activity degree of browsing, bookkeeping, shopping cart addition, and payment can be clearly described. At the same time, according to the statistics of different time spans (week and month) and different operation behaviors, the changing trend of user behavior over time is described to

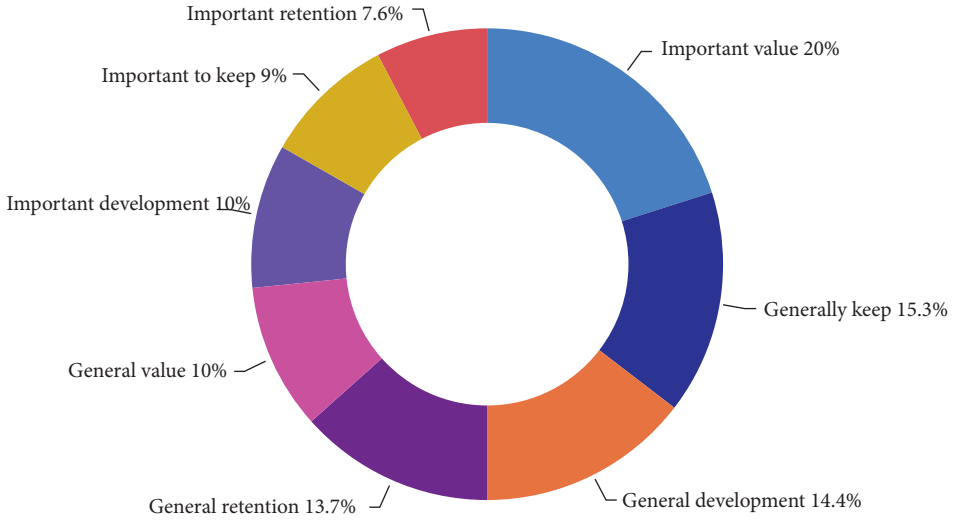


Figure 6: User classification results of this paper.

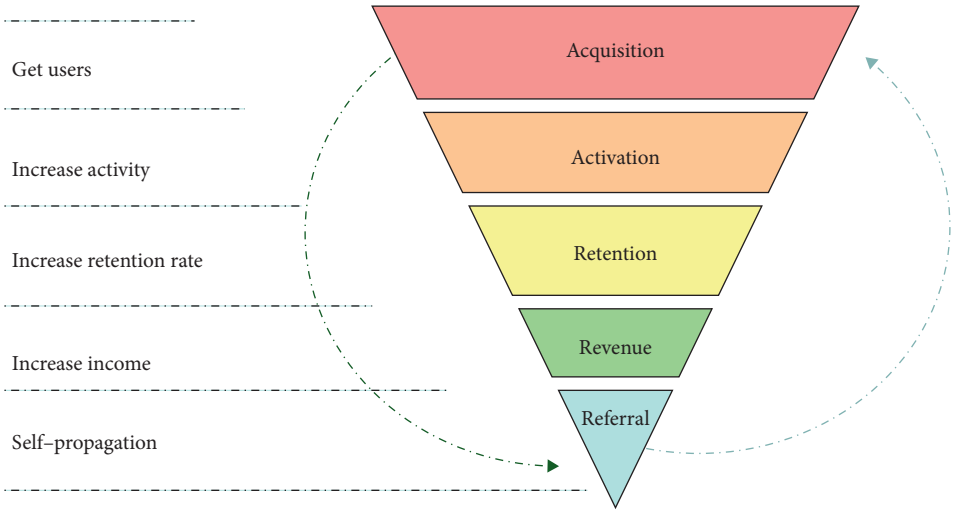


Figure 7: AARRR model.

reflect the shopping habits and commodity preference developed by users over time so that e-commerce platforms can timely grasp the changes of individual user behavior for accurate marketing.

For all users of the e-commerce platform, the number of user operations is described in different time dimensions (day, week, and month), and the group activity degree of all users of the e-commerce platform is described from a different time granularity. Figure 8 shows the change trend of the group user behavior in unit time. In particular, the analysis of each time period of a day can analyze the activity degree of group users in each time period of the day. At the same time, the e-commerce platform can summarize the daily times of user operation behaviors to display the monthly trend of user behavior changes. By analyzing the changes of group users' operation behaviors, the e-commerce platform can find the periodicity of group users' behaviors and make timely decisions to improve the platform revenue.

Based on the composite sequential visualization method, the browsing times, collection times, and adding times to the shopping cart and the payment times of goods on the e-commerce platform are described according to different time dimensions (day, week, and month). By summarizing the changes of a commodity in each period of time, the daily changes of a commodity can be obtained. By summarizing the sales trend of a commodity in a month, the potential factors affecting the sales of a commodity could be obtained by comparing the payment situation with other user operations. At the same time, we could sort the daily payment times and shopping cart times of a certain product to get the daily sales champion and the most popular product, and according to the product ranking, we could change the promotion efforts accordingly to increase revenue.

6.2. Multidimensional Spatial-Temporal Visual Analysis Method Based on Log Mining. In this paper, a multidimensional spatial-temporal visual analysis method based on log mining

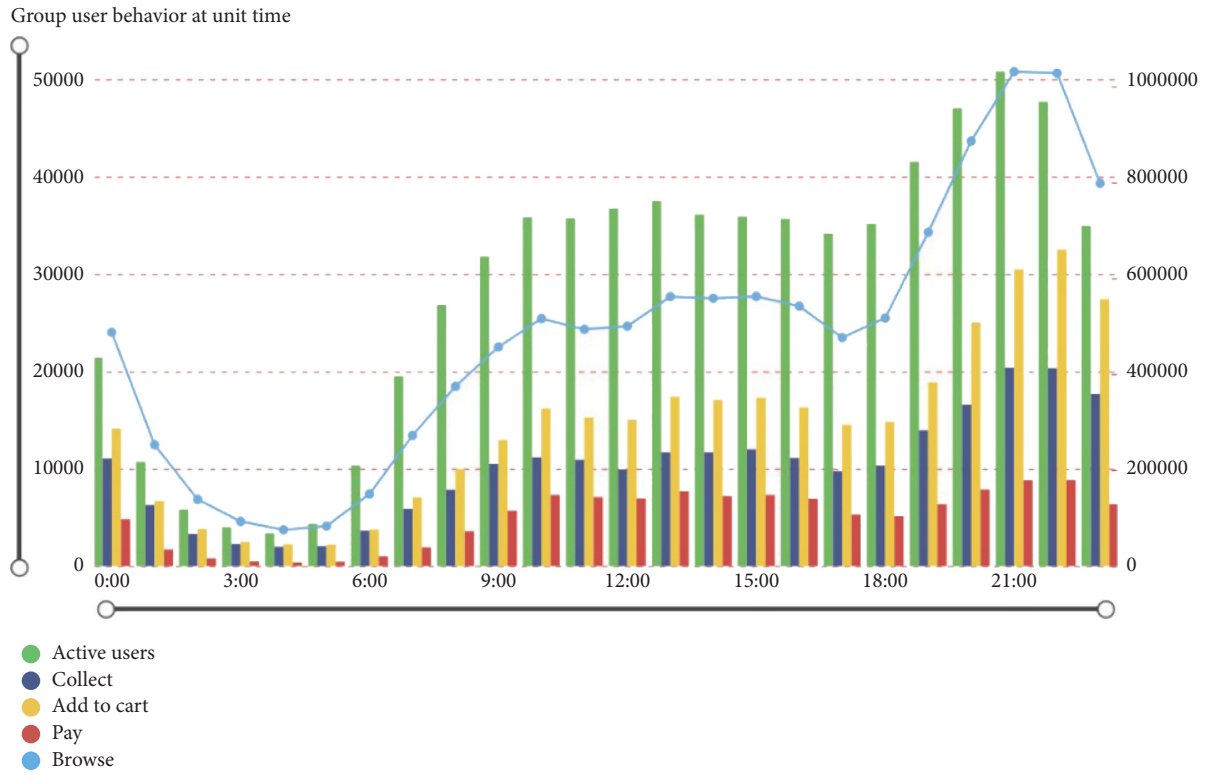


Figure 8: Change trend of group user behavior.

is designed and implemented. These methods mainly include a chord diagram visualization method to show in detail the direction in which goods are flowing. The map visualization method is used mainly to show the sales of goods and user distribution relationship. The funnel diagram visualization method, through funnel analysis, is used to gradually show the loss and transformation of user transaction process. The radar map visualization method is used to show the multidimensional display of individual user behavior and group user behavior.

6.2.1. Chord Graph Visualization Method Based on Data Flow. A chord diagram is a graphical visualization method that shows the relationship between the data in a matrix, which could show the changes in the flow of commodities between different regions. The flow of goods on e-commerce platforms includes both inbound $P(x_1, x_2, x_3 \cdots x_n)$ and outbound $E(y_1, y_2, y_3 \cdots y_n)$ for each region. The sum of the input and output datasets constitutes a dataset N. The purpose of visualization processing is to display the relationship between the matrix data in a two-dimensional space and transform the multivariable into the radian value of the chord graph using the method of weight ratio. The specific change process is as follows: let the dataset of the national commodities be $S(N_1, N_2, N_3 \cdots N_n)$, namely $N_i = \sum_{k=1}^{n}(x_k + y_k)$. The conversion is as follows: $C[l(N_i), w(N_i)]$. The transformed N data points are arranged along the circumference and radially to form a circle as the chord graph of data flow. This data point set is called the node dataset of the chord graph dataset S, denoted as $S'(l_i, w_i)$. The transformation formula of its weight N_i into radian value l_i is shown as follows:

$$l_i = 2 \times \left(\frac{N_i}{\sum_{i=1}^{n} N_i} \right) \times \pi \times r, \tag{3}$$

N_i is the original data of each region, π is PI, r is the radius of the circle of the set chord chart, and l_i is the radian value of N_i after transformation.

The weight of node data is used to determine the size ratio of the arc (shown as Figure 9(2)) and the position of the node in the ring. The weight of the source node and the weight of the target node determine the width of the arc (shown as Figure 9(1)). The weight is converted into the arc width ω_i as follows:

$$\omega_i = \frac{a_i}{N_i} l_i, \tag{4}$$

where a_i is the original data of input or output, N_i is the sum of the amount of input and output data in the i^{th} region, and l_i is the radian value of the i^{th} node data. A node of the arc size is the sum of all the line width, arcs at the node are tiled without overlapping, also node can be classified by color, which is the intuitive display matrix data flow. The flow of data between the data nodes can be one-way or two-way. Through this feature, showing the electric business platform changes the flow of goods.

6.2.2. A Funnel Chart Visualization Method Based on User Dynamics. Funnel diagram visualization method is mainly

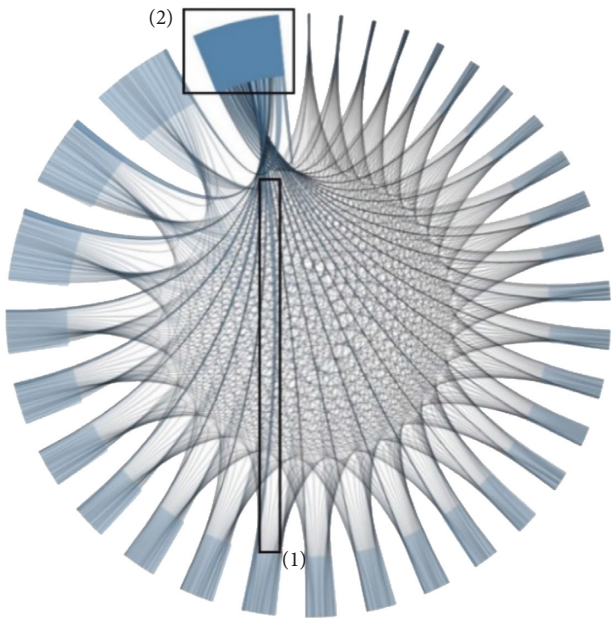


Figure 9: Visualization of chord diagrams based on data flow.

used to analyze the loss and conversion of users at various stages in the shopping process. A shopping operation usually goes through seven steps, which are browsing, collecting, adding shopping cart, checking order, submitting order, selecting a payment method, and completing the payment. There is a certain loss of users in each step. For the loss and conversion of users between each step, the input multivariate data is univariate data, and the purpose of visual processing is to map the dynamic changes of multiple univariates to a two-dimensional space. The use of the funnel diagram method could intuitively express the churn and conversion of users. Let the set of users in each step be $P(x_1, x_2, x_3, \cdots x_7)$. The conversion rate formula of each step is shown as follows:

$$r_i = \frac{x_i}{x_{i-1}} \times 100\%. \tag{5}$$

Here, x_{i-1} is the flow through the upper layer, x_i is the flow reaching the layer, and r_i is the conversion rate between each step. As for the funnel plot, the conversion rate of each step is mapped to the funnel plot by color mapping with the percentage of conversion as the weight, and the funnel plot could well display the loss and conversion of each step.

As shown in Figure 10, browsing is 42.22%, favorites are 27.52%, add shopping cart is 12.51%, check order is 8.63%, submit order is 4.96%, select payment method is 2.62%, and complete payment is 1.54%. From the figure, we see that there is an obvious trend of reducing from browsing to collecting, which may be caused by inaccurate product description, inconsistent product price, unclear product picture, and other problems affecting the user conversion rate. Using the funnel analysis to identify the weak step in current business processes can help analysts focus more on the weak links and improve the overall process output, and therefore, improve the revenue.

6.2.3. Radar Map Visualization Method Based on User Behavior. The description of user value should not only be evaluated according to the purchase volume of users but also analyze users' browsing and collecting of different kinds of commodities and realize a comprehensive evaluation of user value by integrating the user behaviors of different dimensions (time, location, and commodity preference). For users, the multivariate data of each dimension of users are univariate data. Thus, the purpose of visualization processing is to map multiple univariate data to two-dimensional space and transform each variable into amplitude value in the radar chart by the polar coordinate method. The specific change process is as follows: let the behavioral dataset of user i be $B(x_1, x_2, x_3 \cdots x_n)$. There is a transformation relation $f[a(x_i), w(x_i)]$. After transformation, n data points fall exactly in the two-dimensional space represented by the polar coordinates, namely the multidimensional data radar graph, and the graphic dataset $B'(a_i, w_i)$ of dataset B is obtained. As the data range represented by each dimension of the radar chart is limited by the graphic scale of each dimension and the setting range of the graphic scale of each dimension that corresponds to the original data is different, the original data x_i needs to be converted into the corresponding amplitude value a_i during the drawing of each dimension, and the transformation formula is given as follows:

$$a_i = \frac{(M - m) \times (x_i - x_{\min})}{(x_{\text{Max}} - x_{\min})} + m, \tag{6}$$

where x_i is the original data of each dimension, x_{min} is the minimum original data of the i^{th} dimension, x_{Max} is the maximum original data of the i^{th} dimension, m is the minimum amplitude of the set radar chart, M is the maximum amplitude of the set radar chart, and a_i is the amplitude of the original data x_i after transformation. When $m = 0$, $M = 1$, for the radar diagram, rays are drawn based on the center of the circle, and rays in each dimension represent an evaluation index. The radar diagram is a unit circle, and the coordinate values of variables in each dimension after normalization are marked on the corresponding rays. For the drawing of the group user radar chart, the amplitude values of each dimension of each user in the user set are added and averaged, i.e., the amplitude values of each dimension of the group user radar chart are obtained. Figure 11 shows the individual user behavior radar chart (Figure 11(a)) and group user behavior radar chart (Figure 11(b)) for a certain product. The individual user behavior radar chart includes six evaluation dimension, such as clicks, collections, the number of added shopping carts, the number of submitted orders, the number of completed payments, and user value. The group user behavior radar chart includes seven evaluation dimensions, such as click quantity, collection quantity, the number of shopping cart added, the number of orders submitted, the number of groups, the number of payments completed, and the number of buybacks.

6.3. Multidimensional Linkage Visual Analysis Method Based on Log Mining. In this paper, a multidimensional linkage visual analysis method based on log mining is designed and

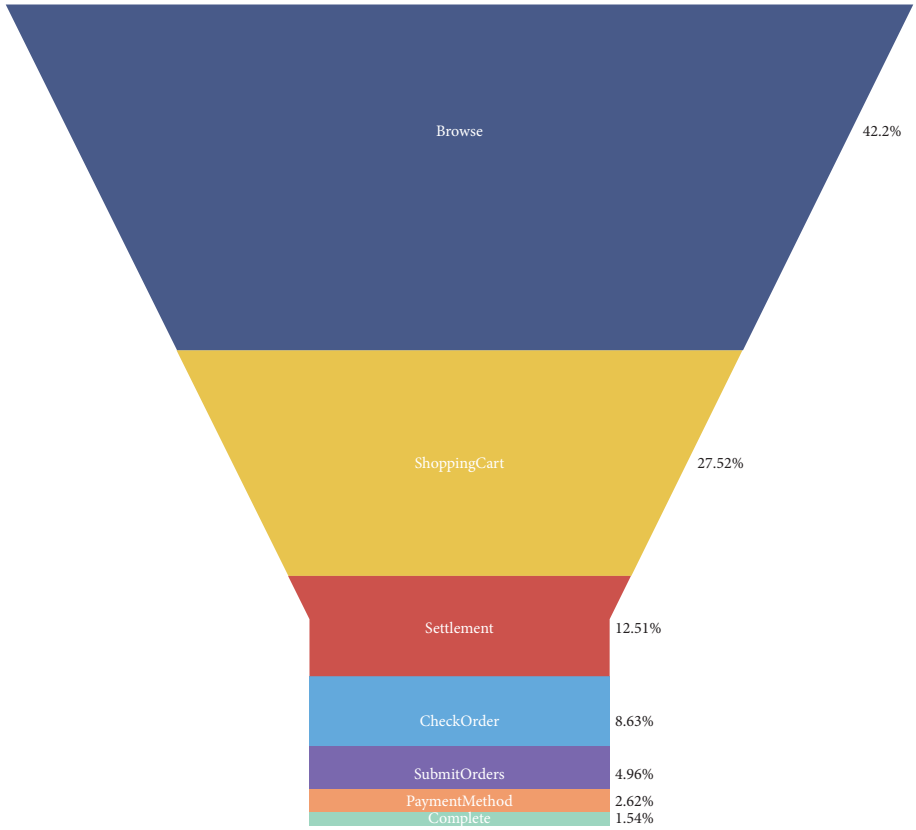


Figure 10: Funnel diagram visualization based on user dynamics.

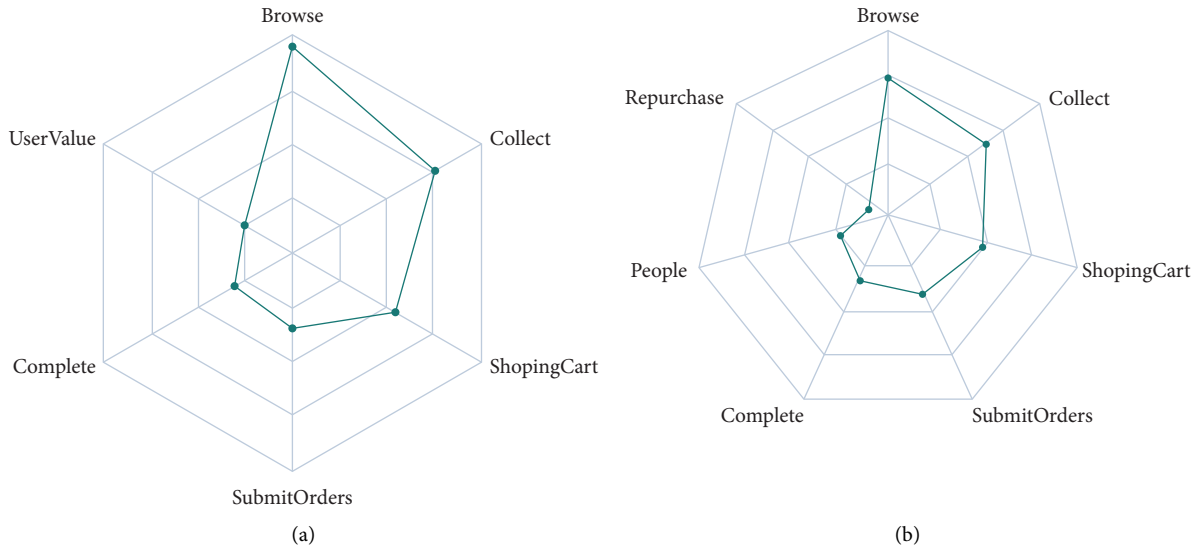


Figure 11: Radar graph visualization based on user behavior. (a) Individual user behavior radar map. (b) Group user behavior radar map.

implemented, which is mainly used to analyze the influencing factors of user behavior changes, analyze the potential laws of increase or decrease of commodity sales, and analyze data changes from multiple dimensions by means of linkage multigraphs centering on the same theme. Faced with the multidimensional analysis of massive user behavior log data, this paper adopts multidimensional cube analysis based on Kylin. The working principle of Apache Kylin is essentially multidimension on-line analysis processing (MOLAP) cube, also known as multidimensional cube analysis. MOLAP is based on a cube called an OLAP Cube. Given a data model, we can aggregate all the dimensions on it. For N dimensions, there are $2N$ possible combinations. For each combination of dimensions, the measures are aggregated, and the results are saved into a materialized view called cuboid. The cuboid of all dimensions combined as a

whole is called cube, as shown in Figure 12. In data multidimensional analysis, the following operations can be performed for a cube: drill (Figure 12(a)), roll (Figure 12(b)), slice (Figure 12(c)), slice (Figure 12(d)), and rotate (Figure 12(e)).

Take Figure 13 multidimensional cube as an example to explain one by one.

(1) Drill: subdivide the aggregated data into smaller data, for example, subdivide data from the second quarter to April, May, and June, and subdivide data from provinces to prefecture-level cities.

(2) Roll: the reverse operation of drilling, i.e., the aggregation of fine-grained data from the high level, for example, the data of April, May, and June are aggregated into the data of the second quarter, and the aggregation of prefecture-level city data into provincial data.

(3) Slice: select a specific dimension for analysis, such as the monthly sales of a specific item.

(4) Dice: select data from a specific dimension for specific analysis, such as analyzing sales changes in a specific quarter throughout the year.

(5) Pivot: select the coordinates of different dimensions for interchange, for example, swap the commodity dimension with the region dimension by rotation.

7. Results and Discussion

In this section, the user behavior log dataset of the Taobao Mall is used as the test data to verify the method proposed in this paper. By the analysis of group user behavior, individual user behavior, and commodity sales, this paper proves the system's ability to solve practical problems and the effectiveness of the visualization method.

7.1. Analysis of Operation Behavior of Group Users

7.1.1. Sequence Diagram of Group User Behavior. Figure 14 shows the time-sharing behavior partition diagram of users on e-commerce websites, which reflects the change trend of all users' behaviors within 24 hours. First of all, the total number of clicks, the total number of favorites, the total number of shopping cart additions, and the total number of payments of all users in each period of time every day are added. From the figure, we could see that the active users of the platform started to increase at morning 8:00, and between 10:00 am and 06:00 pm, the curve integral change is not significant, and the user activity is relatively stable. After 06:00 in the evening, the curve rise is obvious, and a sharp rise in the number of active users reaches the highest at 09:00 pm. The pay times reach the highest throughout the day. After 22:00 hrs., the trend of the curve changed dramatically. The number of active users decreased significantly, and the users reached the lowest point around 3:00 am. The active users showed a rising trend at 5:30 am.

Figure 12 shows the time-sharing behavior heat graph of users on an e-commerce platform. The graph describes the user's activity degree through different shades of color according to the distribution of traffic data to show the change of user behavior in different periods of time every day and intuitively understand the overall data flow of the website through the graph. Figure 15 is time-sharing statistics and superposition of logs of all users, where each square represents the average user behavior operand at that moment. X axis represents time period, Y axis represents user behavior, and from top to bottom are the number of active users, favorites, added shopping carts, submitted orders, and completed payments, respectively. From the perspective of color distribution, from 20:00 to 22:00, users are the most active, and the operation of collecting and adding shopping carts increases significantly. The analysis shows that users like to choose goods from 20:00 to 22:00 in the evening every day.

Figure 16 shows the weekly behavior partition chart of e-commerce platform users, which reflects the change trend of all users' behaviors in a week. First of all, the daily operation times of all users are recorded. A specific date interval is selected. The dates in the interval are marked into blocks by week, and the data in each block are accumulated and added to obtain the weekly user behavior partition graph. It can be observed from the figure that the number of active users in each week is basically unchanged. From Monday to Thursday, the number of users clicking, bookmarking, and adding shopping carts is constantly increasing. On Friday, the overall operation of the user behavior significantly reduced. It is important to note that users pay the amount at its highest level throughout the week. The reason may be that users selected and compared goods from Monday to Thursday and completed payment on Friday. On Saturday, the activity level of users dropped to the lowest, and after Sunday, the activity level of users showed an upward trend.

7.1.2. Group Users Pay Conversion Rates. Figure 17 is a bar chart of the daily payment conversion rate of e-commerce website users, which reflects the daily payment conversion rate change of all users in a certain period of time.

First of all, the daily operation of all users is subdivided into seven types, which are browsing, collecting, adding shopping carts, checking orders, submitting orders, selecting payment choice, and completing payment. Then, the seven data are added to obtain the number of users with different types of operation behaviors. Through funnel analysis, the number of users completing the payment step is compared with the number of users in each step, and the sum of the results is finally obtained to get the daily payment conversion rate of e-commerce platform. We could observe that the users pay conversion rate reached 2.38% on December 12. The pay rate of that day is the highest during the month, the reason may be that the platform's promotion activities on that day activated users' consumption enthusiasm, and, other pay date conversion rates fluctuated, especially on December 1. The pay rate is obviously protruding, and the conversion pay rate is as high as 1.05%. The reason may be the beginning of the Double 12 presale activities, stimulating the users.

Figure 18 is a bar chart of the payment conversion rate of different mobile phone brands, which reflects the change of payment conversion rate of user groups using different

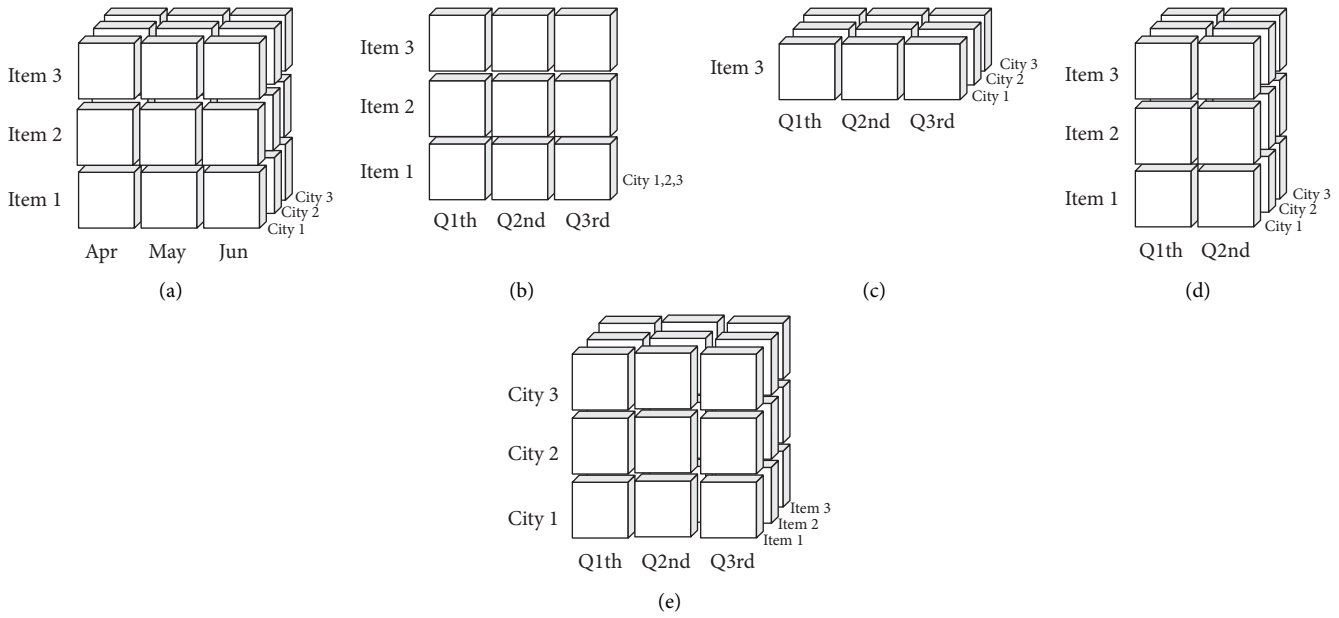


Figure 12: Time-sharing behavior heat diagram of users.

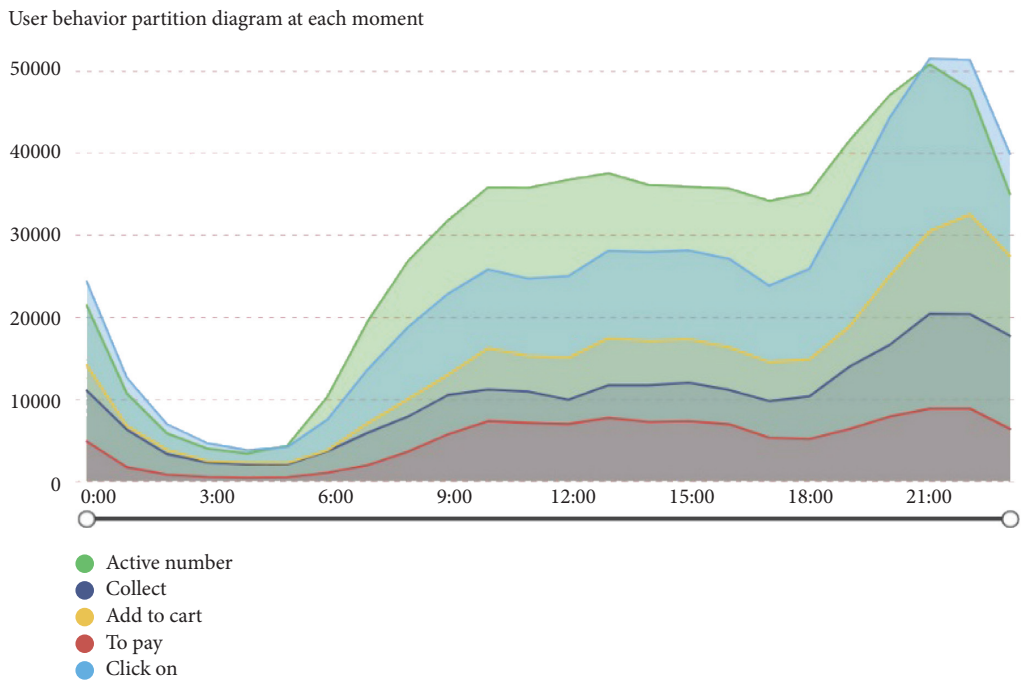


Figure 13: Multidimensional analysis operation of OLAP.

mobile phone brands. Firstly, the user group is subdivided into 7 clusters by the clustering algorithm, namely HUAWEI, Apple, Xiaomi, SAMSUNG, OPPO, Vivo, and Others. Funnel analysis is conducted on the users in the cluster to obtain the payment conversion rate of group users to obtain the payment rate of group users using different mobile phone brands. According to the payment rate of different mobile phone brands, the payment rate is 1.14%, 1.08%, 1.02%, 0.97%, 0.92%, and 0.87, respectively. The figure helps analysts understand the influence of mobile phone brands on payment conversion rate. For mobile phone brands with low payment conversion rate, the reason may be that the e-commerce platform has poor compatibility with this type of mobile phone, and poor user experience can be used as a reference index for platform optimization.

7.1.3. New User Retention Rate. Figure 15 shows the daily turnover rate of new users on e-commerce websites, which reflects the change of retention rate of new users in a period of time.

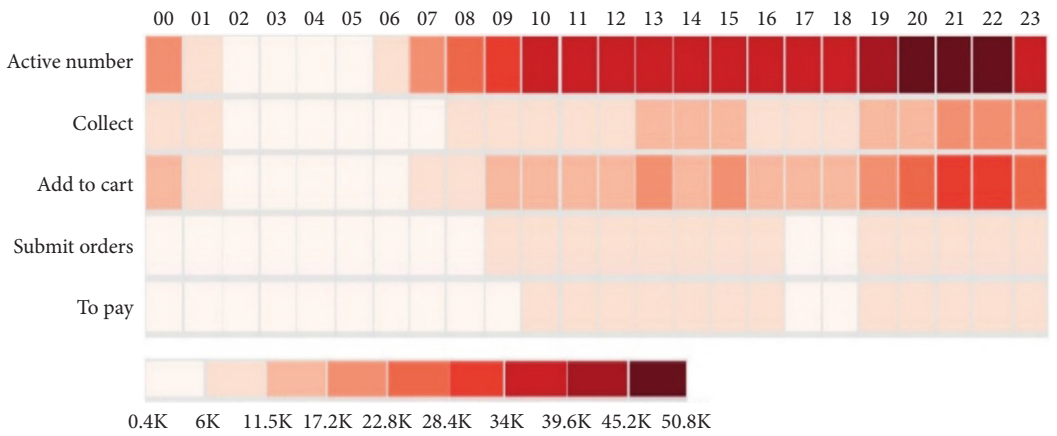


Figure 14: Time-sharing partition diagram of user behavior.

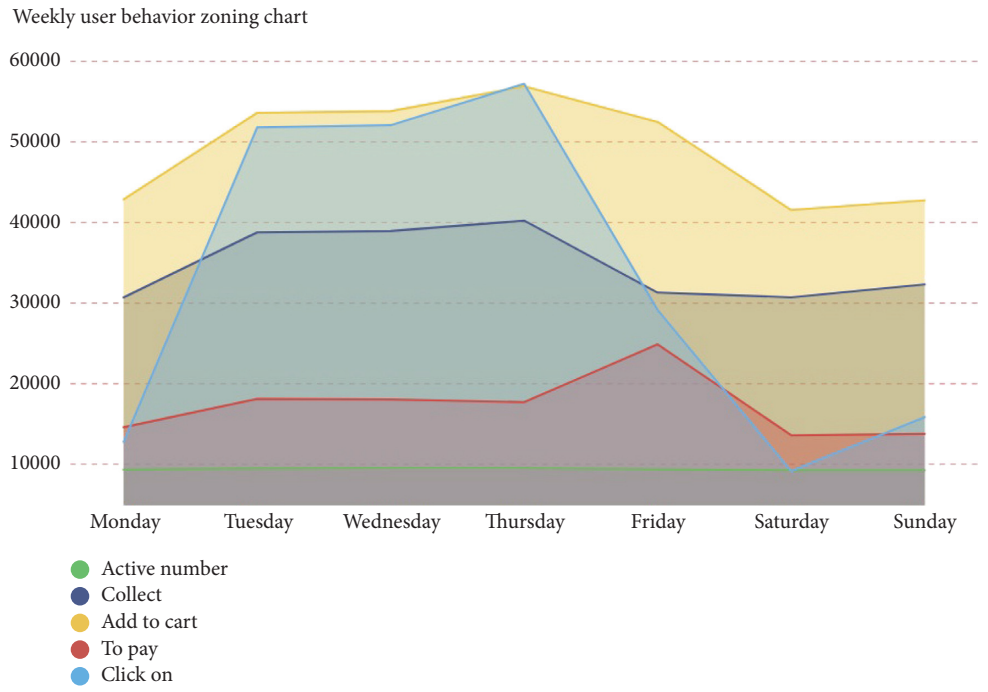


Figure 15: Daily retention filaments.

In this paper, the next-day retention, three-day retention, four-day retention, five-day retention, six-day retention, one-week retention, and half-month retention are set for new users, and the calculation equation is shown as follows:

$$R_i = \frac{x_i}{A} \text{ (i = 2, 3, 4, 5, 6, 7, 15).} \quad (7)$$

Here, R_i represents the retention rate in different time periods, x_i refers to the number of users newly added on the first day who still log in after i days, and A refers to the number of users newly added on the first day. It is worth noting that retention is generally a discrete concept, and users are not required to log in every day within these days. In this way, the retention rate of the next day may be greater than that of the previous day.

Figure 19 shows the silk chart of time-sharing user retention rate. The variation of user retention rate at time-sharing period in Figure 19 can be obtained by drilling down the daily user retention rate in Figure 15. From 30 November to December 1, the user retention rate of the overall shows an upward trend, and the user retention rate presents a fast increasing trend at 12:00 a.m. on November 30. The possible reason is that the new users could be using the spare time after a meal at noon to choose and buy. It is worth noting that the half-month retention rate shows a downward trend before December 1 and an upward trend after December 1, which proves that the promotional activities could increase customer stickiness and improve user retention.

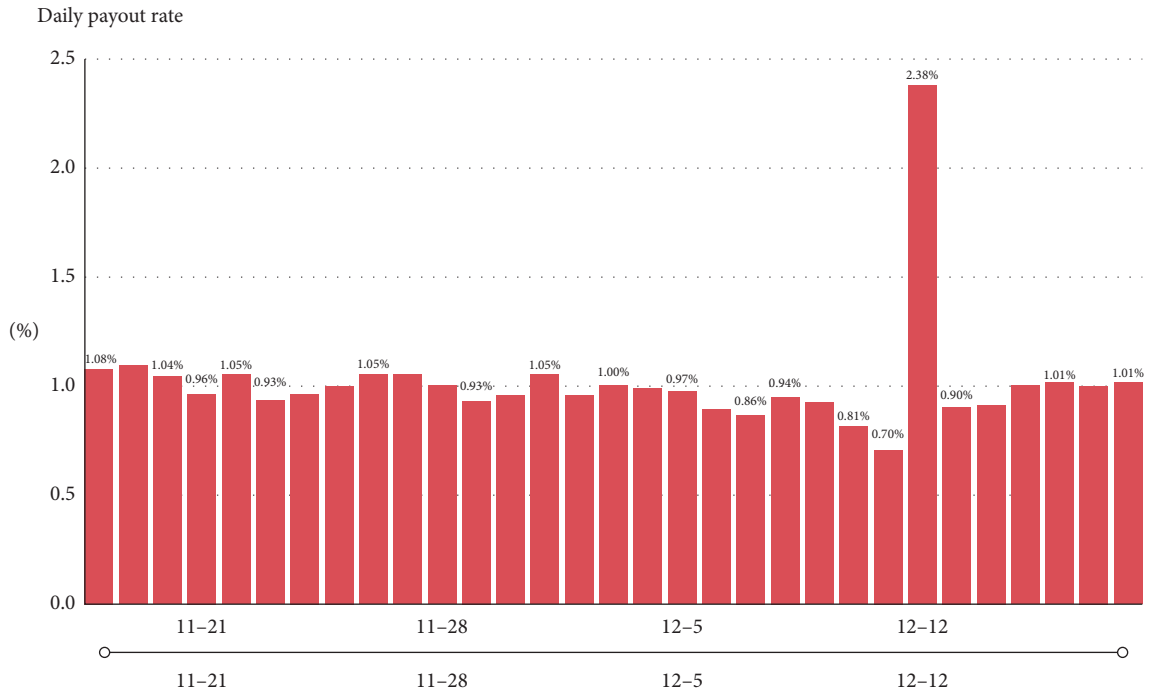


Figure 16: User behavior partition.

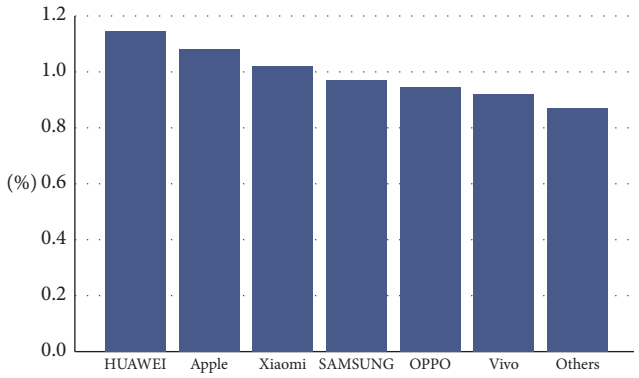


Figure 17: User daily payment rate.

7.1.4. Group User Behavior Radar Map. In this paper, the user value is subdivided by the RFM model, and the user behavior radar map is displayed according to the value of different group customers. Figure 20(a) shows the radar map of the behavior of important value customer groups for purchasing electronic products. The click of this group is 186,754 times, save is 26,325 times, add shopping cart is 20,354 times, and submit orders is 1,865 times. 1,757 payments were made, and 365 repurchases were made. According to the amplitude conversion formula of the radar chart, the amplitude values of this group in clicks, group number, collection, number of shopping carts added, number of orders submitted, number of payments completed, and number of buybacks are 0.589, 0.401, 0.287, 0.2, 0.197, 0.187, and 0.069, respectively. Through the radar map, we can intuitively have a quantitative understanding of group user behavior, help analysts quickly grasp the behavior habits and potential behavior rules of users, formulate relevant marketing strategies in time, and improve the revenue of the platform.

Figure 20(b) is a radar chart comparing the behavior of general developed customer groups and important value customer groups in purchasing electronic products, in which general developed customer groups click 236,754 times, collect 56,325 times, add the shopping cart 17,842 times, submit orders 864 times, pay successfully 755 times, and buy back 47 times. By superimposing the radar maps of the two user groups, the comparison of various dimensions between the groups can be intuitively carried out. The closed graph area formed by each dimension is a multivariate function, which reflects the comprehensive value of the group users, and the area size can be referred to in the formulation of commercial marketing strategies to achieve accurate marketing.

7.2. Analysis of Individual User Behavior

7.2.1. Individual Users Pay Conversion Rates. Figure 21 shows the histogram of individual users' payment conversion rate, which is generated by funnel analysis based on individual users' behavior log data and transaction log data, and it reflects the distribution of individual users' payment conversion rate in the process of shopping. For example, the user with the ID "90241071" has a payment conversion rate of 3.05%, and the user with the ID "90333804" has a conversion rate of 0.2% in the process of shopping. The bar chart clearly shows the user's payment conversion rate, helping analysts find high-quality users and formulating marketing strategies for different users to improve user engagement. The platform could rank users according to the payment conversion rate and conduct cluster analysis, which help

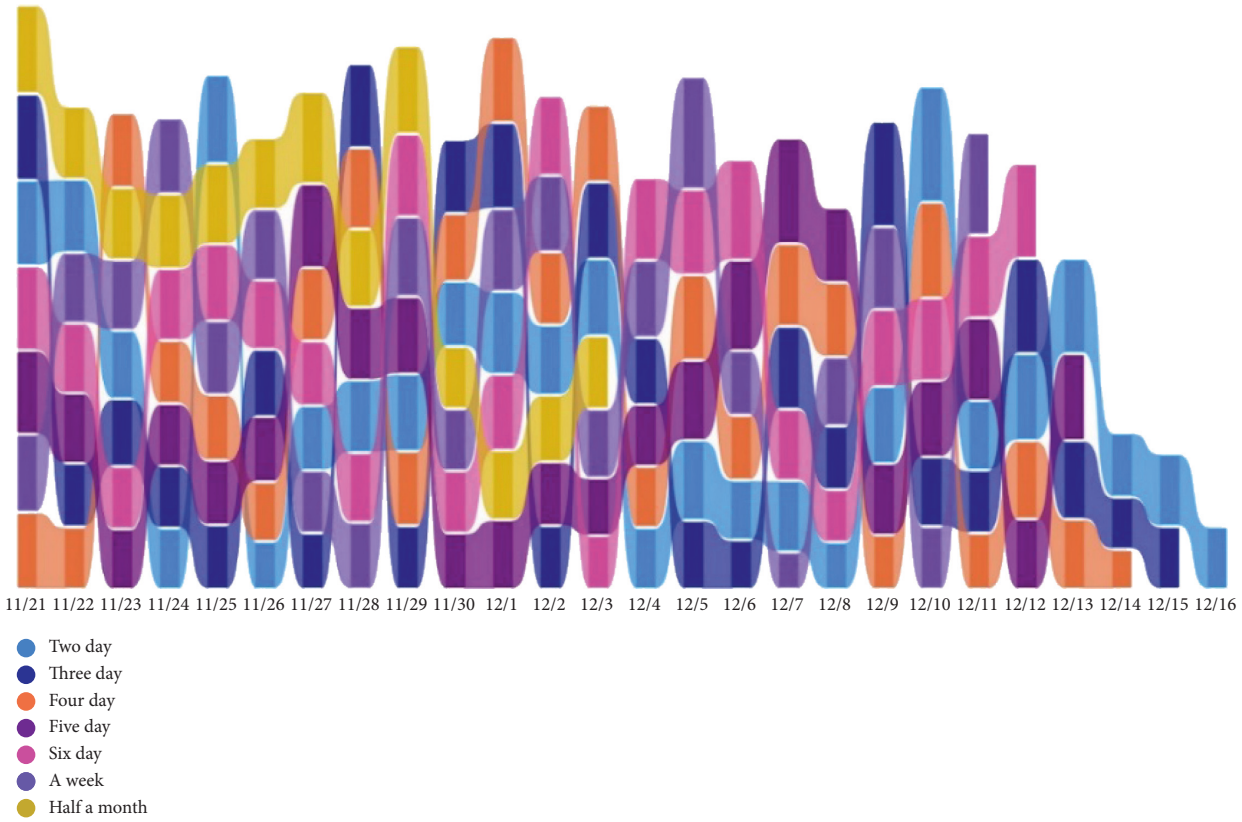


Figure 18: Payment rates of different mobile phone brands.

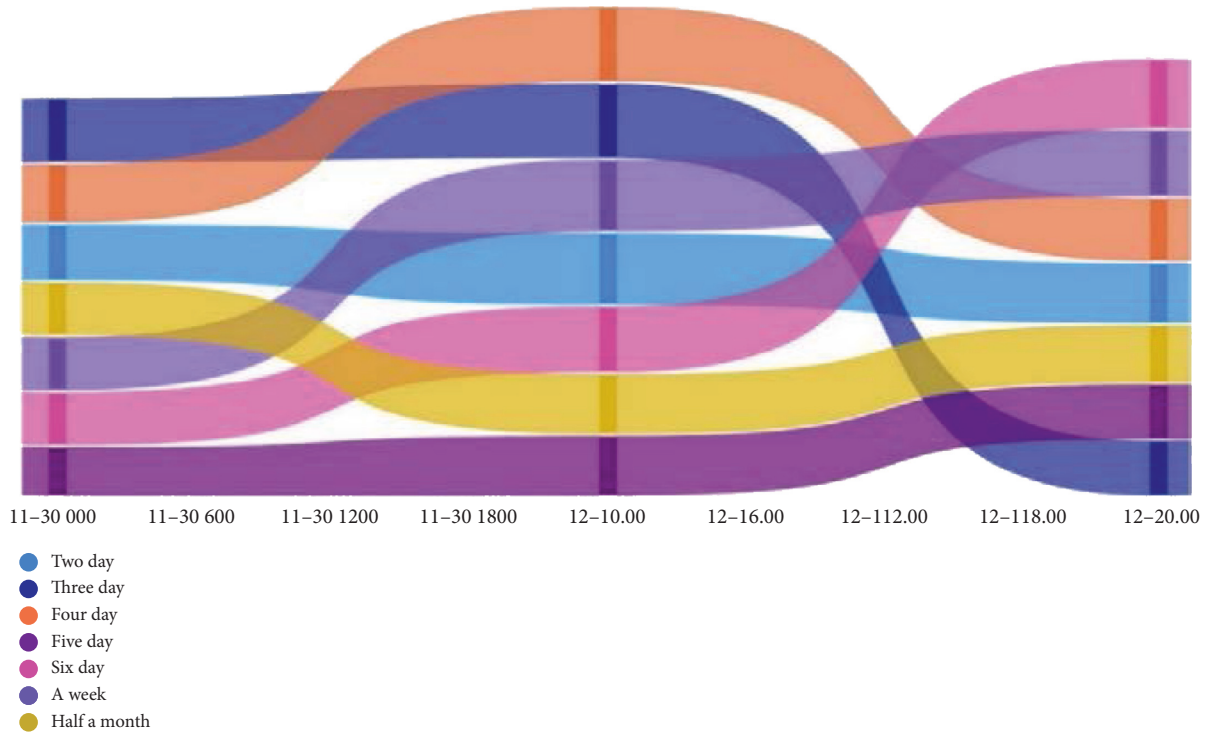


Figure 19: Time sharing user retention rate filamentary chart.

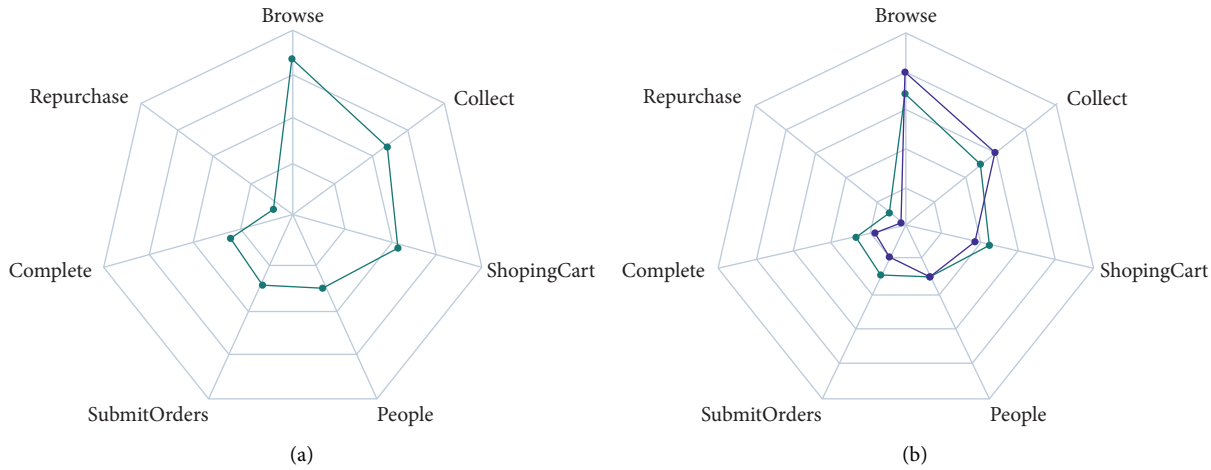


Figure 20: Group user behavior radar map. (a) Group user behavior radar map. (b) Comparison of user behavior among different groups.

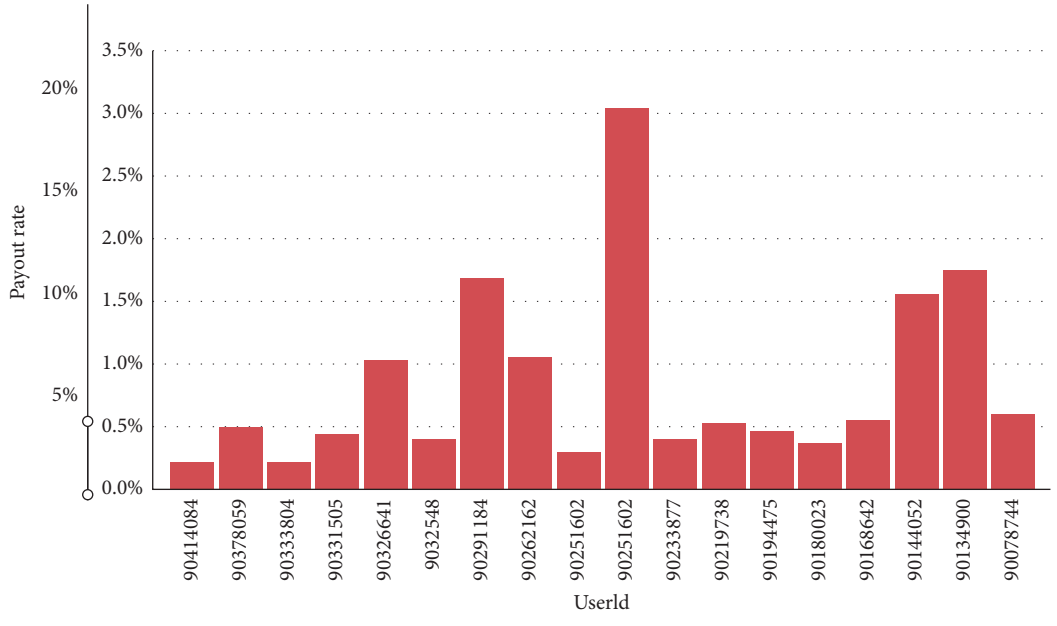


Figure 21: Payment rate of individual users.

analysts to find suitable people for different products to achieve accurate marketing.

According to funnel analysis, the average payment conversion rate is 1.54% if there is a large difference between a user's payment conversion rate and the average payment rate, which indicates that the user has a high loss rate and low user stickiness in the shopping process. However, the single payment conversion rate bar chart cannot find the exact factors affecting the user payment rate. To solve the problem, in Section 7.2.2, this paper adopts the radar chart of individual user behavior to assist the analysis of individual user behavior to make up for the insufficient analysis of the bar chart of single payment conversion rate.

7.2.2. Radar Chart of Individual User Behavior. As shown in Figure 22, the number of operation behaviors of the user with the ID "90333804" in the shopping process is

Figure 22: Individual user behavior radar.

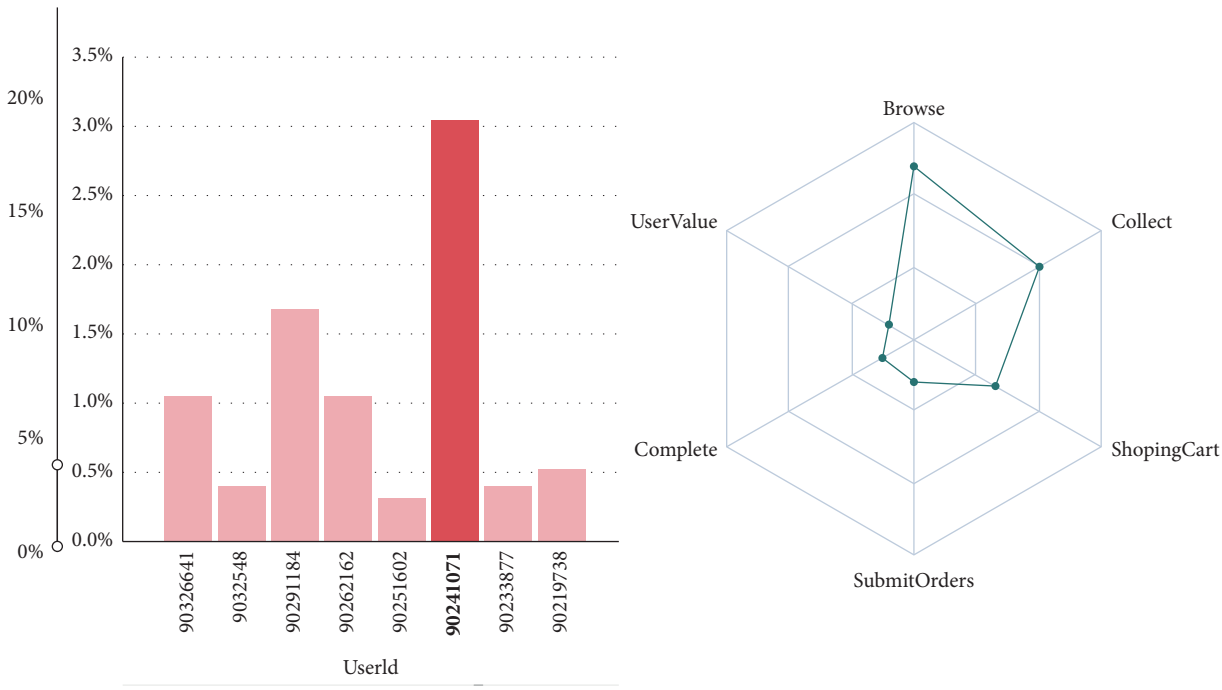


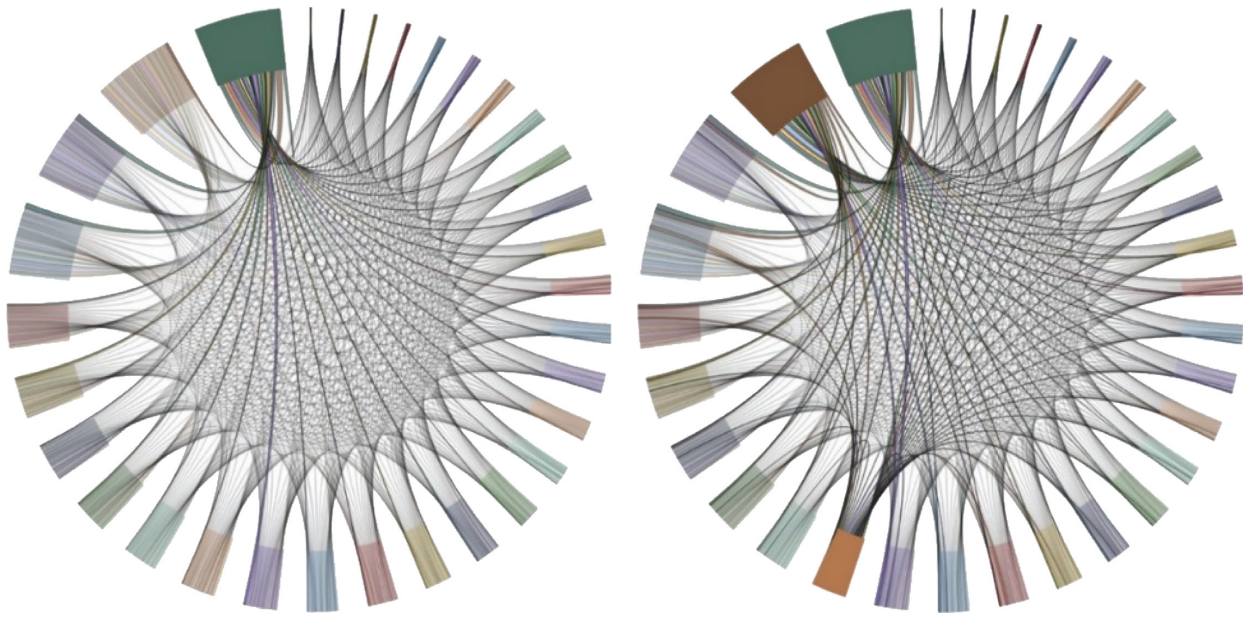
Figure 23: Linkage analysis of individual users.

distributed, including 532 browsing times, 167 bookkeeping times, 102 adding shopping cart times, 5 order submission times, 1 payment completion time, and a user value of 42 points. According to the amplitude transformation formula of radar chart, the amplitude of user "90333804" is 0.587, 0.502, 0.327, 0.159, 0.137, and 0.124 in browsing, bookmarking, adding shopping cart, submitting order, completing payment, and user value dimensions in sequence. Through the individual user radar map, we can have a quantitative understanding of user behavior change intuitively, which helps analysts to have a comprehensive assessment of a user, and adjust marketing strategy to improve the platform revenue. In Section 7.2.3, this paper conducts linkage analysis on individual user radar chart and individual user payment conversion rate. Thus, analysts could select one or more users according to individual user payment rate and can map the amplitude values of each dimensions of one or more users into a radar chart to clearly display the differences between different users.

7.2.3. Linkage Analysis of Individual User Behavior. As shown in Figure 23, the linkage analysis of the payment conversion rate and the behavioral radar chart of user "90241071" is performed. By selecting the payment rate of users, the analyst obtains the behavioral radar chart of individual users at the same time, in which there are 506 browsing times, 165 saving times, 122 adding shopping cart times, 17 order submission times, 15 payment completion times, and a user value of 87 points. According to the amplitude transformation of radar chart, the amplitudes of user "90241071" in browsing, bookmarking, adding shopping cart, submitting order, completing payment, and user value dimension are 0.579, 0.5, 0.364, 0.341, 0.315, and 0.307 in sequence. As for the payment conversion rate, the user's payment conversion rate is as high as 3%, which is much higher than the average payment conversion rate. For individual user behavior radar map, the user has a low loss rate in each step and a high customer stickiness, and the size of the radar map area is the embodiment of the user's comprehensive value. The linkage analysis method is effective to improve the efficiency and accuracy of analysis, and we could also carry out linkage comparative analysis of multiple users at the same time, which could effectively help analysts to have a comprehensive evaluation of user value.

7.3. Product Sales Analysis

7.3.1. Flow Analysis of Goods. By selecting one or more regional nodes in the chord diagram of commodity flow in Figure 24, the sales volume and purchase volume of commodities in one or more regions can be analyzed. This paper selects the commodity flow data from November 28, 2014, to December 28, 2014 for visual analysis, as shown in Figure 24(a). We analyze the commodity sales and purchase volume of "Guangdong Province." Among them, 8,411 commodities were sold from the Guangdong province to the Jiangsu Province, and the Jiangsu province became the province with the largest purchase volume in the Guangdong Province. It is worth noting that the Guangdong province is still the largest province to buy their own goods. The sale amount of goods amounted to 4356. As shown in Figure 24(b), we select Guangdong Province, Jiangsu Province, and Hebei Province for visual analysis of the sales volume and purchase volume. The total



(a) (b)

Figure 24: Chord chart of commodity flow. (a) Chord diagram of commodity flow in a single region. (b) Chord diagram of commodity flow in multiple regions.

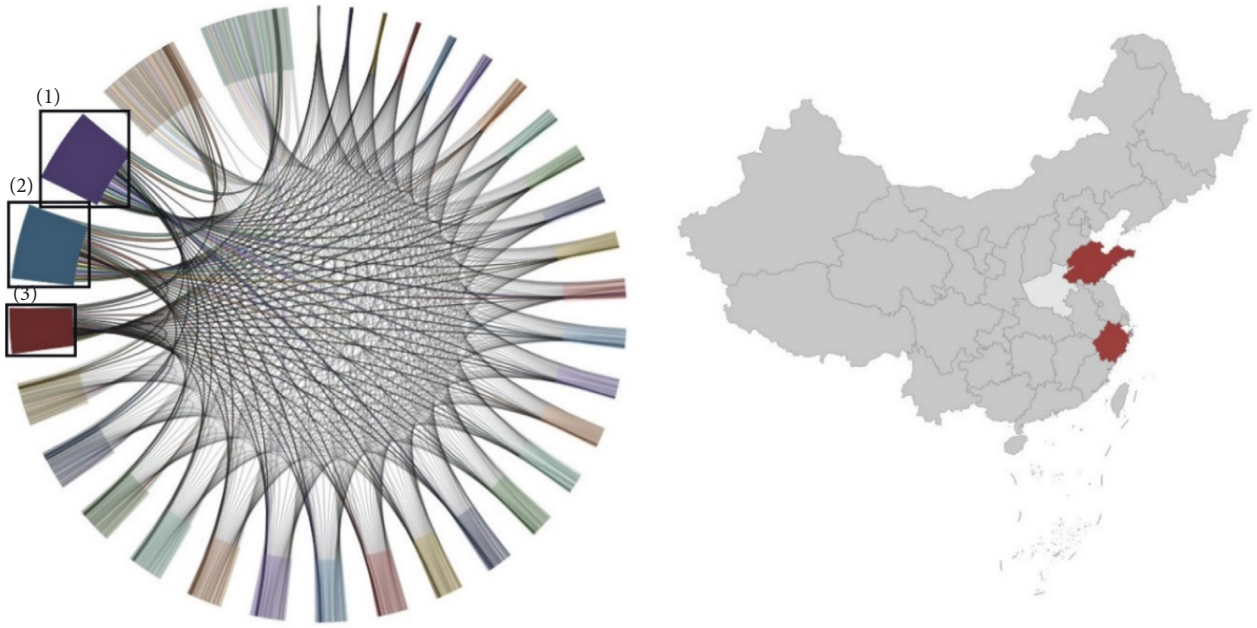


Figure 25: Linkage analysis of commodity sales.

sales volume is 179,512, 176,081, and 131,828, respectively, and the total purchase volume is 10,512, 15,693, and 14,378, respectively.

In this paper, the chord graph visualization method based on data flow is used to show the flow distribution of goods between regions. Through the analysis of regional sales volume and purchase volume, it helps analysts to master the flow of goods to discover the influence and potential law of regional sales volume and purchase volume. In view of chord graph, there are shortcomings in analyzing the influence of user activity degree on commodity sales. In Section 7.3.2, this paper analyzes the relationship between the user activity degree and commodity sales using the linkage analysis of the chord graph of commodity flow and user activity degree of color mapping.

7.3.2. Linkage Analysis of Commodity Sales. By clicking the interest node in the chord diagram of commodity flow in Figure 25, the commodity flow situation in this region can be

viewed, and the distribution of user activity degree in this region could also be displayed in the linkage.

In Figure 25, Shandong Province (Figure 25(1)), Zhejiang Province (Figure 25(2)), and Henan Province (Figure 25(3)) are selected from the chord diagram of commodity flow, and the color mapping method is adopted to display the user activity on the map to gain an insight into the potential value rules through the linkage. Among them, the total sales volume of the Shandong province was 167,008, and the sales volume of the agricultural products was the highest. The sales volume of the commodities throughout the country was mainly sold to the Zhejiang province, Jiangsu Province, and Fujian Province. The total sales volume of commodities in the Zhejiang Province was 166,485, and the sales volume of electronic products was the highest, among which 4,435 were sold to the Jiangsu Province and 4,671 to the Guangdong Province. According to the chord diagram of commodity flow, the total sales volume of the Henan province is 140,160 pieces, which is smaller than that of Shandong and Zhejiang provinces. The linkage analysis of user activity by color mapping shows that the user activity degree of Shandong and Zhejiang provinces is higher, while that of the Henan province is lower. Through experimental verification, it is concluded that the user activity degree and the product sales are mutually promoting, and the product sales also increase in the region with higher user activity degree.

8. Conclusion

This paper studies the visual analysis of e-commerce user behavior. Using log mining technology and visual analysis methods, such as composite timing visualization, chord diagram visualization, map visualization, and radar diagram visualization, this paper designs and implements a data warehouse system that can store and analyze massive data and visually display the results. Through the log data analysis based on the user behavior of the Taobao Mall, the effectiveness of the *K*-means clustering algorithm based on user behavior segmentation, the multidimensional spatial-temporal visual analysis method based on log mining, and the multidimensional linkage visual analysis method based on log mining proposed in this paper are verified. Complete the analysis of individual and group users' comprehensive value, behavior time rule, shopping conversion rate, and other behavioral characteristics. Realize the effective mining and multidimensional visual analysis of the e-commerce user behavior, which provides data support for the internal decision-making of enterprises and provides data support for the realization of accurate marketing.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the State Key Program of National Natural Science Foundation of China under Grant no. 41901297.

References

[1] R. C. Basole and W. B. Rouse, "Complexity of service value networks: conceptualization and empirical investigation," *IBM Systems Journal*, vol. 47, no. 1, pp. 53–70, 2008.

[2] H. Chen, fnm Chiang, and fnm Storey, "Business intelligence and analytics: from big data to big impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165–1188, 2012.

[3] R. Lengler and M. J. Eppler, "Towards a periodic table of visualization methods for management," in *Proceedings of the IASTED Proceeding of the Conference on Graphics and Visualization in Engineering Figure 11. Architecture for data visualization*, Florida, USA, January 2007.

[4] K. Zhang, "Using visual languages in management," *Journal of Visual Languages & Computing*, vol. 23, no. 6, pp. 340–343, 2012.

[5] M. Ankerst, "Circle segments: a technique for visually exploring large multidimensional data sets," in *Proceedings of the Proc. Visualization '96*, Hot Topic Session, San Francisco, CA, January 1996.

[6] D. A. Keim, "Pixel-oriented visualization techniques for exploring very large databases," *Journal of Computational & Graphical Statistics*, vol. 5, p. 58–77, 1996.

[7] P. Kotler, *Marketing Management – Analysis, Planning and Control*, Prentice-Hall, New Jersey, USA, 4th ed edition, 1980.

[8] M. J. Crotft, *Market Segmentation: A Step-Bystep Guide to Profitable New Business*, Routledge, England, UK, 1994.

[9] M. Sarstedt and E. Mooi, *A Concise Guide to Market Research*, pp. 273–324, Springer Texts, Berlin, Germany, 2014.

[10] A. Weinstein, *Market Segmentation: Using Demographics, Psychographics and Other Niche Marketing Techniques to Predict and Model Customer Behavior*, Probus Pub. Co, Chicago, 1994.

[11] V. Venugopal and W. Baets, "Neural networks and statistical techniques in marketing research," *Marketing Intelligence & Planning*, vol. 12, no. 7, pp. 30–38, 1994.

[12] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[13] J. McQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, L. M. Le Cam and J. Newman, Eds., vol. 1, pp. 281–297, University of California Press, California, USA, 1967.

[14] Y.-H. Hu and T.-W. Yeh, "Discovering valuable frequent patterns based on rfm analysis without customer identification information," *Knowledge-Based Systems*, vol. 61, pp. 76–88, 2014.

[15] U. Kaymak, "Fuzzy target selection using rfm variables," in *IFSA World Congress and 20th NAFIPS International Conference,* vol. volume 2, pp. 1038–1043, IEEE, 2001.

[16] S. M. S. Hosseini, A. Maleki, and M. R. Gholamian, "Cluster analysis using data mining approach to develop crm methodology to assess the customer loyalty," *Expert Systems with Applications*, vol. 37, no. 7, pp. 5259–5264, 2010.

[17] J.-T. Wei, M.-C. Lee, H.-K. Chen, and H.-H. Wu, "Customer relationship management in the hairdressing industry: an application of data mining techniques," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7513–7518, 2013.

[18] S. H. Han, S. X. Lu, and S. C. H. Leung, "Segmentation of telecom customers based on customer value by decision tree model," *Expert Systems with Applications*, vol. 39, no. 4, pp. 3964–3973, 2012.

[19] S.-Y. Kim, T.-S. Jung, E.-H. Suh, and H.-S. Hwang, "Customer segmentation and strategy development based on customer lifetime value: a case study," *Expert Systems with Applications*, vol. 31, no. 1, pp. 101–107, 2006.

[20] A. Z. Ravasan and T. Mansouri, "A fuzzy anp based weighted rfm model for customer segmentation in auto insurance sector," in *Intelligent Systems: Concepts, Methodologies, Tools, and Applications*, pp. 1050–1067, IGI Global, USA, 2018.

[21] H.-C. Chang and H.-P. Tsai, "Group rfm analysis as a novel framework to discover better customer consumption behavior," *Expert Systems with Applications*, vol. 38, no. 12, pp. 14499–14513, 2011.

[22] J.-T. Wei, S.-Y. Lin, C.-C. Weng, and H.-H. Wu, "A case study of applying LRFM model in market segmentation of a children's dental clinic," *Expert Systems with Applications*, vol. 39, no. 5, pp. 5529–5533, 2012.

[23] E. R. Tufte, *Beautiful evidence*, Vol. volume 1, Graphics Press Cheshire, CT, 2006.

[24] Z. Liu, J. Stasko, and T. Sullivan, "SellTrend: inter-attribute visual analysis of temporal transaction data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1025–1032, 2009.

[25] R. Chang, M. Ghoniem, R. Kosara et al., "Visualization of categorical, time-varying data from financial transactions," in *Proceedings of the IEEE VAST*, pp. 155–162, IEEE, Sacramento, CA, USA, November 2007.

[26] M. C. Hao, J. Ladisch, U. Dayal, M. Hsu, and A. Krug, *Visual mining of e-customer behavior using pixel bar charts*, in *Proceedings of the Proc. ACM KDD/2001*, pp. 1–7, San Francisco, CA, USA, 2001.

[27] D. A. Keim, M. C. Hao, U. Dayal, and M. Lyons, "Value-cell bar charts for visualizing large transaction data sets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 4, pp. 822–833, 2007.

[28] W. Chen, Z. Huang, F. Wu, M. Zhu, H. Guan, and R. Maciejewski, "VAUD: a visual analysis approach for exploring spatio-temporal urban data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 9, pp. 2636–2648, 2018.

[29] H. Hochheiser and B. Shneiderman, "Dynamic query tools for time series data sets: timebox widgets for interactive exploration," *Information Visualization*, vol. 3, no. 1, pp. 1–18, 2004.

[30] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. S.. Lifelines, "Using visualization to enhance navigation and analysis of patient records," *AMIA Symposium*, vol. 76, 1998.

[31] M. Hao, D. A. Keim, U. Dayal, D. Oelke, and C. Tremblay, "Density displays for data stream monitoring," *Computer Graphics Forum*, vol. 27, no. 3, pp. 895–902, 2008.

[32] D. F. Jerding and J. T. Stasko, "The information mural: a technique for displaying and navigating large information spaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 4, no. 3, pp. 257–271, 1998.

[33] M. R. Naeem, H. Naeem, M. Aamir, W. Ali, and W. A. Abro, "A multi-level process mining framework for correlating and clustering of biomedical activities using event logs," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 3, pp. 393–401, 2017.

[34] H. Naeem, B. Guo, M. R. Naeem, and D. Vasan, "Visual malware classification using local and global malicious pattern," *Journal of Computers*, no. 6, pp. 73–83, 2019.

[35] L. Qi, C. Hu, X. Zhang et al., "Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4159–4167, 2020.

[36] H. Kou, H. Liu, Y. Duan et al., "Building trust/distrust relationships on signed social service network through privacy-aware link prediction process," *Applied Soft Computing*, vol. 100, Article ID 106942, 2021.

[37] L. Qi, X. Wang, X. Xu, W. Dou, and S. Li, "Privacy-aware cross-platform service recommendation based on enhanced locality-sensitive hashing," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1145–1153, 2020.

[38] F. Wang, H. Zhu, G. Srivastava, S. Li, M. R. Khosravi, and L. Qi, "Robust collaborative filtering recommendation with user-item-trust records," *IEEE Transactions on Computational Social Systems*, 2021.

[39] https://tianchi.aliyun.com/dataset.