Journal of Electronic Research and Application, 2025, Volume 9, Issue 3
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Application of Big Data Technology in User Behavior Analysis of E-commerce Platforms

Yanzhao Jia*

Amazon Services Co., Ltd., Bellevue 98004, Washington, USA

**Author to whom correspondence should be addressed.*

Abstract: With the rapid development of the Internet and e-commerce, e-commerce platforms have accumulated huge amounts of user behavior data. The emergence of big data technology provides a powerful means for in-depth analysis of these data and insight into user behavior patterns and preferences. This paper elaborates on the application of big data technology in the analysis of user behavior on e-commerce platforms, including the technical methods of data collection, storage, processing and analysis, as well as the specific applications in the construction of user profiles, precision marketing, personalized recommendation, user retention and churn analysis, etc., and discusses the challenges and countermeasures faced in the application. Through the study of actual cases, it demonstrates the remarkable effectiveness of big data technology in enhancing the competitiveness of e-commerce platforms and user experience.

Keywords: Big data technology; E-commerce platform; User behavior analysis

Online publication: 29 May, 2025

1. Introduction

The rapid development of e-commerce has enabled major platforms to have a large user base and massive transaction data. Understanding user behavior is crucial for e-commerce platforms, which helps platforms to optimize operation strategies, improve user satisfaction and loyalty, and increase sales. Traditional data processing methods are not enough to deal with such large-scale and complex data, while big data technology, with its powerful data processing capabilities, can conduct in-depth analysis of user behaviors on e-commerce platforms and mine out valuable information to provide strong support for platform decision-making.

2. Overview of big data technologies

2.1. Definition and characteristics of big data

Big data refers to a collection of data that cannot be captured, managed, and processed by conventional software tools within a certain time frame, and it is a massive, high-growth, and diverse information asset that requires

new processing modes to have stronger decision-making, insight discovery, and process optimization capabilities. Big data has the characteristics of "4V," i.e., Volume, Velocity, Variety, and Value. In e-commerce platforms, the amount of data generated every day, such as user browsing records, transaction data, evaluation information, etc., is huge, and the data is generated at a fast speed, and the data types cover various forms, such as text, images, and numbers, etc., and at the same time, these data contain great commercial value [1].

2.2. Big data technology systems

Big data technologies cover a range of technologies such as data collection, storage, processing, and analysis. Data collection technologies include web crawlers, log harvesting, etc., which are used to obtain data from various sources. For data storage, there are distributed file systems (e.g., Hadoop Distributed File System, HDFS) and non-relational databases (e.g., MongoDB, Redis), etc., which can efficiently store massive data. Data processing technologies mainly include MapReduce, Spark, etc., which can perform parallel computing on large-scale data and improve processing efficiency. Data analysis technologies include data mining, machine learning, deep learning, etc., through which valuable information and patterns can be extracted from data [2].

3. Collection and storage of user behavior data on e-commerce platforms

3.1. Sources of data on user behavior

Table 1. User behavior

Data type	Specific data values	Analyzed meaning
Page view data	User A browses the "Mobile Phones" category page, stays there for 120 seconds, and clicks on the "iPhone 15" product detail page	Understand user interests and optimize page layout and recommendation algorithms
Click on the link data	User B clicks on the "Limited Time Offer" banner and is redirected to the promotion page	Evaluate advertising effectiveness and optimize marketing strategies
Search Keyword Data	User C searched for "wireless headphones," found 50 results, clicked on product #3	Analyze user needs and optimize search algorithms and keyword recommendations
Purchase of recorded data	User D purchased the mobile phone "Huawei Mate 60," the order amount is 5999 RMB, the payment method is Alipay	Analyze users' spending power to optimize product pricing and promotional strategies
Order information data	User E's order consists of "iPhone 15" and "AirPods Pro," and the order status is "Shipped"	Analyze user purchase mix and optimize product bundling strategy
Payment method data	User F used a credit card to pay for 2999 RMB, the payment was successful	Analyze user payment preferences and optimize payment channels and offers
Product evaluation data	User G rated "iPhone 15" 5 out of 5 stars, saying "Great photos, great battery life"	Understand user satisfaction with products and improve product quality and description
User recommendation data	Suggestion by User H: "Would like to add more color options"	Understanding user requirements to optimize product design and supply chain management
Complaint data	User I complained: "Logistics speed is too slow, promised 3 days delivery, but the actual use of 7 days"	Improvement of logistics services to enhance user satisfaction
Social media sharing of data	User J shared his experience with the iPhone 15 on Weibo, captioning, "The photo effect is amazing!"	Analyze user word-of-mouth communication to optimize social media marketing strategies
Social media interaction data	User K liked the "AirPods Pro" promotion posted by the platform's official account on Instagram	Evaluate the effectiveness of social media campaigns and optimize user interaction strategies
User behavioral trajectory data	User L's behavioral trajectory: Home → Search "wireless headphones" → Browse product details → Add to cart → Place order	Analyze user decision paths to optimize user experience and conversion rates
User profile data	User M: gender male, age 28, city Beijing, high consumption level, prefers electronic products	Building user profiles for accurate marketing and personalized recommendations

The server of an e-commerce platform records every user action on the website, such as page browsing, clicking on links, searching for keywords, and so on. Website log data records the user's behavioral trajectory in detail and is an important data source for user behavior analysis. Transaction data includes users' purchase records, order information, payment methods, and so on. Transaction data reflects the user's actual purchasing behavior and is of great significance to the analysis of the user's consumption habits and purchasing preferences. Information such as users' evaluations and scoring of goods, as well as their suggestions and complaints, can help platforms understand users' satisfaction with goods and services, as well as users' needs and expectations. Some e-commerce platforms cooperate with social media platforms to obtain platform-related content shared by users on social media, such as product recommendations and usage experiences. Social media data can provide a richer perspective for user behavior analysis [3].

3.2. Data collection techniques

With the help of programs written in specific programming languages and based on pre-set rules, web crawlers can automatically traverse web pages and capture the required data. In the highly competitive e-commerce field, e-commerce platforms use crawler technology to efficiently obtain competitors' product details, including style, material, features, and other information, while closely tracking their price dynamics to provide a strong reference for their pricing strategy. In addition, it can also collect user evaluations of its products on other social platforms and product review sites to gain a comprehensive understanding of user feedback. Log collection tools, such as Flume and Logstash, play a key role on the server side; Flume is based on a reliable data streaming mechanism that captures log data dispersed across server nodes in real time and transfers it to a centralized storage area in an orderly manner; Logstash, on the other hand, has powerful data processing and conversion capabilities, and can clean, format, and pre-process the data while collecting log data. Logstash, on the other hand, has powerful data processing and conversion capabilities. The two work together to integrate server log data, laying a solid foundation for subsequent in-depth analysis of user behavioral trajectories and operation frequency. Buried point technology cleverly embeds code snippets in the platform's web page code or application program, so that when a user performs specific operations such as clicking on a product detail button, adding a shopping cart, or completing a page browse, the code is triggered to automatically send data to the server that contains detailed information about the operation time, user identification, and content of the operation, to accurately capture every micro-behavior of the user, and to help e-commerce platforms to analyze user behavioral preferences and provide a basis for the subsequent deep analysis of user behavioral trajectories. This helps e-commerce platforms to deeply analyze users' behavioral preferences and provide accurate data support for personalized services and product optimization [4].

3.3. Data storage methods

HDFS is a distributed file system that stores data on multiple nodes with high fault tolerance and high scalability. In e-commerce platforms, a large amount of user behavior data can be stored on HDFS, which is convenient for subsequent data processing and analysis. MongoDB is suitable for storing unstructured and semi-structured data, such as user comments, log records, etc. It has a flexible data model and efficient read and write performance. It has a flexible data model and efficient read/write performance, which can meet the needs of e-commerce platforms for massive data storage and fast querying. Redis is an in-memory database with extremely high read/write speed. In e-commerce platforms, it is commonly used to store user session information, popular product data, etc., to improve the response speed of the system [5].

4. Application of big data technology in user behavior analysis

4.1. User profile construction

Collect basic information of users such as age, gender, geography, etc., browsing behavior, purchasing behavior, evaluation behavior, and other multi-source data, and integrate these data. Extract attributes from the integrated data that can reflect the user's characteristics, such as the user's hobbies, consumption ability, purchase frequency, and so on. These features are analyzed through data mining and machine learning algorithms to determine the user's behavioral patterns and preferences. Based on the analysis results, a user portrait model is constructed. The user profile model can be represented in the form of labels, each label corresponding to a characteristic of the user. For example, if a user often buys high-end electronic products, it can be labelled as "high-end electronic product enthusiast" and "high spending power" [6].

4.2. Precision marketing

Using big data technology to segment users, users are divided into different groups based on their age, gender, geography, consumption behavior, and other characteristics. For example, users can be divided into high consumption groups, low consumption groups, new user groups, old user groups, and so on. Develop personalized marketing strategies for different user groups. For high consumption groups, high-end products and exclusive offers can be pushed; for new user groups, newbie packages and guided marketing activities can be provided; for old user groups, loyalty and exclusive member benefits can be launched. Through precise marketing, the targeting and effect of marketing activities can be improved, and marketing costs can be reduced. Evaluate the effect of marketing activities by analyzing the user's behavioral data before and after the marketing activities, such as click rate, conversion rate, and purchase amount. According to the evaluation results, timely adjust the marketing strategy, optimize the marketing activities, and improve the marketing ROI [7].

4.3. Personalized recommendations

Collaborative filtering-based algorithms find other users with similar interests to the target user based on the user's behavioral data, and then recommend the commodities preferred by these similar users to the target user. For example, if user A and user B both purchased commodity X and commodity Y, and user A also purchased commodity Z, then commodity Z can be recommended to user B. Content-based recommendation algorithms recommend commodities to users that are similar to commodities they have previously browsed or purchased based on the attributes of the commodities and the user's preferences. For example, if a user frequently browses sports goods, then sports goods of other brands or related sports accessories can be recommended. Deep learning recommendation algorithms use deep learning models to analyze user behavioral data and learn the relationship between user interest patterns and commodities to achieve more accurate recommendations. For example, neural network models are used to model the user's browsing history, purchase records, and other data to predict the user's preference for different commodities. Real-time recommendation leverages the real-time processing capability of big data technology to recommend relevant commodities for users in real time based on their real-time behavior (e.g., commodities being browsed, current search keywords, etc.) [8]. Real-time recommendations can improve the timeliness and relevance of recommendations and enhance user experience. Recommendation effect optimization optimizes the recommendation algorithm and improves the accuracy and quality of recommendations by continuously collecting user feedback data on the recommendation results, such as whether to click on the recommended products and whether to buy the recommended products. At the same

time, the recommendation strategy is adjusted according to different scenarios and user needs, such as launching corresponding recommendation programs during holidays and promotional activities [9].

4.4. User retention and churn analysis

Commonly used retention metrics include next-day retention rate, seven-day retention rate, monthly retention rate, and so on. By analyzing these indicators, it is possible to understand the retention of users in different periods. Use big data technology to analyze factors that affect user retention, such as product quality, service level, user experience, etc. For example, by analyzing users' evaluation data and complaint records, key issues affecting user satisfaction can be identified, and timely improvements can be made. Based on the results of retention analysis, develop targeted retention strategies. For new users, you can improve the first-time experience of users through newbie guidance and preferential activities to promote user retention; for old users, you can introduce personalized services and benefits to enhance user stickiness. Identify possible lost users by setting certain rules and indicators, such as not logging in for a long time and not purchasing goods. Use big data technology to analyze the behavioral data of churned users to find out the reasons for user churn. For example, by comparing the behavioral characteristics of churned users with those of retained users, it is found that churned users may have encountered payment problems, out-of-stock products, etc., during the purchase process [10]. Based on the reasons for churn, develop appropriate recovery strategies. For users lost due to price factors, we can provide exclusive promotional activities; for users lost due to service problems, we can promptly resolve user complaints and provide compensation measures. Through the recovery of lost users, reduce the losses caused by user loss [11].

5. Case studies

5.1. Amazon's big data applications

Amazon is one of the world's largest e-commerce companies, and it is a leader in the application of big data technology. Amazon builds a detailed user profile by collecting data on users' browsing history, purchase records, search behavior, and so on. Based on user profiles, Amazon provides personalized recommendation services for users, with a recommendation accuracy rate of over 60%. In terms of precision marketing, Amazon sends personalized emails and push notifications to users based on their behavioral characteristics and interests, increasing the conversion rate of marketing activities. At the same time, Amazon uses big data technology to analyze user retention and churn, and continuously optimizes the user experience and service quality of the website to maintain a high user retention rate [12].

5.2. Alibaba's big data practices

Alibaba owns several e-commerce platforms, such as Taobao and Tmall. Alibaba uses big data technology to analyze huge amounts of user behavior data to achieve precision marketing and personalized recommendations. For example, in the annual "Double 11" shopping carnival, Alibaba uses big data technology to segment users and develop personalized marketing strategies for different user groups, attracting a large number of users to participate in the event. In terms of user retention, Alibaba analyzes user behavioral data to identify the pain points and needs of users in the shopping process, and continuously optimizes the functions and services of the platform to improve user satisfaction and loyalty. In addition, Alibaba also applies big data technology to supply chain management, logistics, and distribution, improving the operational efficiency of the entire e-commerce ecosystem [13].

6. Challenges and countermeasures for the application of big data technologies

6.1. Data quality issues

Data collected by e-commerce platforms may have problems such as noise, missing values, and duplicate values, which affect the accuracy and reliability of data analysis. Data cleaning technology can be used to denoise the data, fill in missing values, remove duplicate values, and perform other processing. At the same time, a data quality monitoring mechanism is established to monitor data quality in real time and to identify and resolve data quality problems promptly [14].

6.2. Data security and privacy protection

User behavior data contains a large amount of personal information, such as name, ID card number, bank card number, etc., and data security and privacy protection face serious challenges. Once the data is leaked, it will bring huge losses to users and damage the reputation of e-commerce platforms. The application of data security technologies, such as data encryption, access control, and identity authentication, can be strengthened to ensure data security. At the same time, strict data privacy protection policies should be formulated to clarify the scope of data use and authority, and to obtain explicit authorization from users when collecting and using their data [15].

6.3. Shortage of skilled personnel

The application of big data technology requires talents with professional knowledge and skills, including data scientists and big data engineers. At present, there is a shortage of such professionals in the market, which restricts the wide application of big data technology in e-commerce platforms. It is necessary to strengthen the cooperation between universities and enterprises in the training of big data talents, offer relevant professional courses, and cultivate big data professionals who meet the market demand. At the same time, enterprises should strengthen the training of internal staff to improve their ability to apply big data technology.

6.4. Algorithmic complexity and interpretability

Some big data analytics algorithms, such as deep learning algorithms, have a high level of complexity, making it difficult to understand and explain their decision-making process. This affects the credibility and application effectiveness of the algorithms to a certain extent. By researching and developing interpretable algorithms, the decision-making process of the algorithms can be understood and explained. At the same time, when applying complex algorithms, the results of the algorithms are presented to the user intuitively by combining visualization techniques to improve the interpretability and transparency of the algorithms.

7. Conclusion

Big data technology has a wide range of application prospects and important practical significance in user behavior analysis of e-commerce platforms. Through the collection, storage, processing, and analysis of user behavior data, e-commerce platforms can build accurate user profiles, achieve precision marketing and personalized recommendations, improve user retention and reduce user turnover, thus enhancing the competitiveness of the platform and user experience. However, in the process of applying big data technology, it also faces challenges such as data quality, data security, and a shortage of technical talent. By taking corresponding countermeasures, such as improving data quality, strengthening data security protection, and cultivating professional and technical talents, these challenges can be effectively dealt with to promote the in-depth application of big data technology

in the analysis of user behavior on e-commerce platforms, and to provide strong support for the development of the e-commerce industry. In the future, with the continuous development and innovation of big data technology, its application in the field of e-commerce will be more extensive and in-depth, bringing more opportunities and changes to the development of e-commerce.

Disclosure statement

The author declares no conflict of interest.

References

[1] Zhao L, 2024, Research on Marketing Strategy of E-commerce Platform based on Big Data Technology. Old Brand Marketing, (02): 33–35.

[2] You Y, 2023, Application of Big Data Technology in e-commerce Platform. China Quality Miles, (01): 57–59.

[3] Wang Y, 2018, The Use of Big Data Technology in O2O E-commerce. Mall Modernisation, (19): 28–29.

[4] Cheng H, Wu Y, 2018, Exploration of E-commerce Ethical Construction under the Background of Big Data Technology. Journal of Hunan University of Technology (Social Science Edition), 23(01): 124–128.

[5] Xu H, 2017, Application of Big Data Technology in E-commerce Platform and Enterprise. Computer Knowledge and Technology, 13(33): 279–280.

[6] Zhi C, Liu J, Wang Y, 2024, Big Data Swarm Intelligence Optimization Algorithm Application in the Intelligent Management of an E-Commerce Logistics Warehouse. Journal of Cases on Information Technology, 26(1): 1548–7717.

[7] Guo Y, Lin J, Zhuang W, 2024, An Evolutionary Game-based Regulatory Path for Algorithmic Price Discrimination in E-Commerce Platforms. Mathematics, 12(17): 2774. https://doi.org/10.3390/math12172774

[8] Zheng X, 2024, Analysis of E-commerce Enterprises based on Big Data. Modern Business, (24): 185–188. https://doi.org/10.14097/j.cnki.5392/2024.24.041

[9] Lin C, 2024, Application of Big Data Technology in E-commerce Data Analysis. Quality and Market, (11): 72–74.

[10] Tang B, 2024, Research on Big Data-driven E-commerce Marketing Strategy Innovation. Modern Business Research, (18): 119–121.

[11] Zhang X, 2024, Research on the Marketing Path Optimization of E-commerce under the Big Data Perspective. Business Theory in China, 33(16): 35–38.

[12] Fan S, Zhan M, Yao Y, 2024, The Application of Big Data Technology in Agricultural Products E-commerce. Grain in Heilongjiang Province, (05): 88–90.

[13] Chen W, 2024, Research on the Optimization of E-commerce Enterprise Management Mode in the Era of Big Data. Administrative Business Assets and Finance, (08): 36–38.

[14] Jiang Y, 2024, Discuss the Application of Big Data Technology in E-commerce Data Analysis. China Transportation and Storage, (04): 109–110.

[15] Liu X, Zhang S, 2024, Construction of E-commerce Information Management System based on Big Data Technology. Information and Computers (Theoretical Edition), 36(03): 102–104.