Received: 9 Jul. 2025 | Revised: 30 Jul. 2025 | Accepted: 6 Aug. 2025 | Published online: 25 Aug. 2025

RESEARCH ARTICLE

Journal of Global Humanities and Social Sciences
2025,Vol. 6(5)212-218
DOI: 10.61360/BoniGHSS252018640508

Application of Big Data in E-commerce User Behavior Analysis

Qian Deng[1,*]
[1] *Geely University of China, China*

Abstract: In the context of white-hot e-commerce competition and increasingly fragmented user demand, the traditional experience-driven operation model has been difficult to meet the precise and real-time business needs. The rise of big data technology provides a key path to crack this problem. By integrating multi-dimensional behavioral data such as user browsing, searching, purchasing, evaluation, etc., and combining machine learning and real-time analysis technology, it builds up a decision-making system covering intelligent recommendation, user life cycle management, supply chain optimization and other scenarios. The purpose of this paper is to systematically explore the core application of big data in e-commerce user behavior analysis: on the one hand, it reveals how to improve user conversion and retention through data-driven refined operation, on the other hand, it analyzes its practical value in the supply chain to reduce costs and increase efficiency, and commercial decision-making intelligence, etc., so as to provide theoretical support and practical references for e-commerce enterprises to build a competitive advantage in data.

Keywords: user behavior modeling, intelligent algorithm, real-time data analysis, supply chain optimization, data-driven decision-making

Introduction

Under the wave of digitalization, the e-commerce industry is experiencing a profound change from traffic competition to value mining. As a “digital mine” for business decision-making, the scale of user behavior data is growing exponentially and its dimensions are becoming more and more complex, and the traditional analysis methods have been difficult to capture the dynamically changing user needs. The breakthrough of big data technology has provided e-commerce enterprises with a powerful tool to extract value from massive data and gain insight into the deeper intentions of users. This article focuses on the innovative application of big data in the analysis of e-commerce user behavior, and discusses how it can reconfigure the relationship of “human-cargo-field” through intelligent recommendation, user lifecycle management, supply chain optimization and other scenarios, helping enterprises achieve precise operation, efficiency improvement and experience upgrade, and providing a practical paradigm for the industry's digital transformation.

1. Importance of Big Data in E-commerce User Behavior Analysis

1.1 Commercial value driving

The commercial value driving role of big data in e-commerce user behavior analysis is reflected in its subversive innovation and efficiency leap in business model through the deep mining of user interaction data. The whole chain of user behavior data from browsing, clicking, purchasing to payment has become the core assets of e-commerce platforms - these data can not only quantify user preferences (e.g., through the click heat map to identify the area of high concern for goods), consumption ability (e.g., based on the historical order amount to divide the user tier) and social influence (e.g., analyzing the fission level triggered by the act of sharing), but also quantify user's behavior. (e.g. analyzing the fission

Corresponding Author: Qian Deng
Geely University of China, China

level triggered by sharing behavior), but also predict future behavior through machine learning models (e.g. predicting the repurchase cycle of users using LSTM neural network). When data-driven decision-making replaces empiricism, e-commerce platforms are able to achieve precision marketing (a case study shows that personalized recommendations increase conversion rates by 27%), dynamic pricing (real-time price adjustments based on supply and demand relationships increase GMV by 15%) and supply chain optimization (improved demand forecast accuracy increases inventory turnover by 40%), and ultimately build a closed-loop business ecosystem of "data-insight-action-feedback". Finally, it builds a closed-loop business ecosystem of “data-insight-action-feedback”, which significantly improves the user life cycle value (CLV) while reducing the cost of customer acquisition (Si, 2025).

1.2 Technology enabling advantages

Big data technology for e-commerce user behavior analysis gives unprecedented enabling advantages, the core of which lies in the technical architecture innovation and algorithm breakthroughs to achieve the exponential release of data value. At the data processing level, the combination of distributed computing frameworks (e.g., Spark, Flink) and columnar storage (Parquet) enables e-commerce platforms to process millions of user behavior events (e.g., clicks, swipes, and length of stay) in real time per second, and to synchronize real-time analysis and offline modeling through the flow and batch integrated architecture; at the algorithmic level, the deep learning models (e.g., Transformer, graph neural network, etc.) can be used for the analysis of user behavior in real time. At the level of algorithmic application, deep learning models (e.g., Transformer, graph neural network) can tap into the complex correlation patterns of user behaviors (e.g., cross-category purchase sequences, influence propagation in social networks), while reinforcement learning can dynamically optimize the interaction strategies (e.g., real-time strategy adjustment in personalized recommendation systems).

2. The Deep Impact of Big Data on E-commerce

2.1 User level

Big data is profoundly reshaping the user's behavioral patterns and consumption experience in e-commerce, forming a new type of human-computer interaction driven by data. In terms of behavioral transparency, every click, browsing time, product comparison and other micro-behaviors of users are captured in real time, and through clustering analysis and sequence pattern mining, e-commerce platforms can accurately identify the potential needs of users (e.g., the trajectory of searching behaviors of pregnancy and baby products in the three months before purchasing), and even prejudge the changes in life stages (e.g., inferring the needs of housewarming through the pattern of furniture purchasing); in terms of the decision-making path, the intelligent recommender system enabled by Big Data has taken over users' decision-making process. In terms of decision-making path, big data-enabled intelligent recommendation system has taken the lead in user decision-making - research shows that 68% of e-commerce users will directly click on the recommended products to complete the purchase, and the traditional "search-comparison" mode is gradually replaced by "algorithm-guided-instant gratification" (Su & Xu, 2025).

2.2 Enterprise level

Big data is driving e-commerce enterprises to transform from experience-driven to data intelligence-driven, reconfiguring their core competitive elements and operational logic. At the strategic level, the data center has become the new infrastructure of the enterprise, through the integration of user behavior data dispersed in the transaction, logistics, customer service and other systems to form a unified view of data assets to support the whole chain of decision-making from market forecasting to product innovation (such as ZARA's 72-hour rapid replenishment mechanism based on sales data and social media trends); at the organizational level, the data team from the back-office support department jumped up to become the core business engine, giving rise to the “data

engine". At the organizational level, the data team has risen from a back-office support department to a core business engine, giving rise to new positions such as "data product manager" and "algorithmic operation", and promoting changes in cross-departmental collaboration (e.g., Ali's "data BP" system realizes the in-depth coupling of the technical team and the business line); at the efficiency level, the automated marketing system (MA) has increased the ROI of the marketing campaign by more than 40% through the user subgroups and touchpoints management, and the intelligent supply chain based on the demand prediction optimizes the inventory turnover to 2.3 times that of the traditional model. to 2.3 times of the traditional model (Wei & Wang, 2019).

3. Systematic Countermeasures: Solutions for the Whole Chain from Technology to Organization

3.1 Construction of data governance system

The value release of e-commerce user behavior analysis driven by big data highly depends on a perfect data governance system. This system needs to cover the entire life cycle of data collection, storage, processing and application, and through the three core modules of standardization, quality control and bloodline tracking, it can solve the stubborn problems of "data silos", "dirty data" and "untraceability". Take the construction of Alibaba's "data center" as an example: at the standardization level, Ali has formulated the "Data Asset Classification Standard", which divides user behavior data into 3 categories and 21 subcategories (e.g., subdividing "clicks" into "clicks on the product detail page" and "clicks on the advertisement space"), and unifies the definitions and calculation calibers of more than 400 business fields, eliminating cross-sectoral data ambiguities; in the area of quality control, the self-researched "DataWorks" is used as a tool to help users to understand and analyze data. In terms of quality control, its self-developed "DataWorks" platform integrates more than 300 checking rules, which can automatically identify and repair missing values and outliers (e.g., marking orders with a purchase amount of more than a million dollars in a single day as pending verification), thus increasing the data accuracy rate to 99.2%; at the level of lineage tracking, the "Atlas" metadata management system realizes the visualization of the whole chain from the original logs to the analysis reports. When the effect of a marketing activity is abnormal, the operation personnel can locate the problematic nodes in the data processing chain within 30 seconds (e.g., an ETL task omits to process the "add purchase without purchase" user label). With the support of this system, Ali is able to integrate user behavior data scattered in Taobao, Tmall, Alipay and other systems into "OneData" assets to support core business scenarios such as thousands of recommendations and intelligent replenishment in the supply chain (Chen & Ma, 2024).

3.2 Privacy compliance innovation path

In the big data-driven analysis of e-commerce user behavior, privacy compliance innovation has become a core proposition for the sustainable development of enterprises. The path needs to build a closed-loop system around technological innovation, management mechanism and ecological collaboration in three aspects: at the technical level, the use of federal learning, differential privacy and other cutting-edge technologies to achieve the data "available but not visible", for example, the Ant Group through the privacy computing platform to achieve cross-organizational user behavior analysis, the original data is always retained locally, and only the exchange of Encrypted model parameters are exchanged, which not only protects the joint modeling needs of partners (e.g., banks and logistics companies), but also ensures that intermediate computation results can't be parsed in reverse through homomorphic encryption; on the management mechanism, a compliance framework covering the whole life cycle of data is established, such as Jingdong's "Privacy Compliance Center", which integrates the user authorization management, data desensitization rules, access audit logs, and so on. For example, Jingdong has built a "privacy

compliance center", integrating user authorization management, data desensitization rules, access audit logs and other modules into a unified platform, and when users browse commodities, the system automatically triggers the anonymization process, replacing sensitive information such as device ID and geographic location with dynamic tokens, and at the same time, recording the operator, time, and direction of the data flow to satisfy the requirements of the GDPR for "data traceability"; eco-collaboration In terms of eco-collaboration, a trust network is built through standard contract terms and third-party certification. For example, Alibaba's cross-border platform, Speedway, adopts standard contract templates approved by the European Commission in the EU market to clarify the data protection obligations of data receivers (e.g., overseas warehouse service providers), and commissions SGS and other international certification organizations to conduct compliance audits on a regular basis, which, according to the company's financial report for the year 2025, has resulted in a 67% year-on-year decrease in cross-border data transmission complaints. Complaints dropped by 67% year-on-year, while orders from EU companies increased by 41% due to compliance advantages.

3.3 Real-time analysis capability upgrade

Driven by the real-time demand for e-commerce user behavior analysis, real-time analysis capability upgrade has become the core engine for enterprises to seize the user decision-making window and optimize operational efficiency. Its technical evolution needs to focus on the three dimensions of architectural innovation, computation optimization and scene adaptation: at the architectural level, the stream-batch integrated computing framework (e.g., Apache Flink) is replacing the traditional Lambda architecture to achieve a balance between millisecond latency and batch processing accuracy through a unified processing engine, e.g., Pinduoduo adopts Flink+Kafka to build a real-time counting warehouse, which combines the event streams of users' clicks, purchases and payments with offline dimensional tables (e.g., products). For example, Pinduoduo uses Flink+Kafka to build a real-time number warehouse, which dynamically associates the event flow with offline dimension tables (such as product labels and user profiles) to support real-time updating of the "Thousands of People, Thousands of Faces" recommender system, which processed 20 million pieces of behavioral data per second during the Double Eleven promotion, and increased the conversion rate of the recommendation by 18% compared with the traditional offline mode; in terms of computation optimization, the deep integration of memory computation and column storage significantly improves the query efficiency. In terms of computing optimization, the deep integration of memory computing and columnar storage significantly improves query efficiency, for example, Meituan accelerates Hive metadata access through Alluxio, combined with the Parquet columnar storage format, so that the response time of complex analytical queries (such as user cross-category purchase path analysis) is compressed from the minute level to less than 3 seconds; in terms of scenario adaptation, edge computing sinks analytical capabilities to terminal equipment and reduces the pressure on the central node, for example, Xiaomi Yuping deploys lightweight and lightweight computing capabilities to intelligent shelves and other devices. Intelligent shelves deploy lightweight flow processing engine, real-time analysis of user residence length, commodity touch frequency and other micro behavior, when a commodity is detected to be frequently picked up but not purchased, the system immediately triggers dynamic pricing strategies (such as pushing limited-time discount coupons), pilot store data show that the program makes impulse consumption accounted for 25% increase (Wu & Zhou, 2025).

3.4 Organizational capability evolution

In the context of the deep reconstruction of the competitive landscape of e-commerce by big data, organizational capability evolution has become the core proposition for enterprises to transform data potential energy into business kinetic energy. The core is to break the traditional hierarchical barriers

and build a data-driven agile organizational ecosystem, the specific path covers cultural reshaping, role reconstruction and collaboration mechanism innovation in three aspects: cultural level, the need to cultivate a decision-making culture of "talking with data", such as Amazon through the "two pizza team" principle (team size of less than one team). The principle of "two pizza team" (team size does not exceed the number of people who can be fed by two pizzas), data analysis embedded in the smallest business unit, and its "reverse working method" requires that all product decisions must be based on the verification of user behavioral data rather than management intuition, which makes the Prime membership system from the assumption to the ground in just six months, far exceeding the number of members in the market. This culture makes the Prime membership system from hypothesis to implementation in only 6 months, far exceeding the industry average cycle; role reconstruction, data scientists and business personnel skills increasingly fuzzy boundaries, giving rise to "data translator", "algorithm operations officer" and other cross-border positions, such as NetEase YenSelect set up For example, NetEase Yanxuan has set up the position of "user research engineer", which requires members to be proficient in SQL data extraction and have knowledge of consumer psychology, so that they can directly transform the "frequent jumping" behavior in the user's browsing path into the optimization plan of the product detail page; in terms of the innovation of the collaboration mechanism, it is necessary to In terms of innovation of collaboration mechanism, it is necessary to establish a closed loop of two-way feedback between data and business. For example, through the mode of "data war room", Vipshop centralizes the operation, technology and analysis teams, so that when real-time monitoring reveals the abnormal conversion rate of a certain brand of women's apparel, the data analysts can immediately work with the operation to formulate the AB test program (e.g., adjusting the main picture text), and the technology team can synchronize with the development of new function modules, which can be completed within 4 hours. Functional modules, complete the omni-channel push within 4 hours, the mechanism makes its GMV increase by 23% during the promotion period (Ye, 2016).

4. Application of Big Data in E-commerce User Behavior Analysis

4.1 Intelligent recommender system

Big data-driven intelligent recommender system has become the core engine of e-commerce platforms to enhance user experience and commercial efficiency, which realizes the precise recommendation of "thousands of people, thousands of faces" through the integration of multi-dimensional behavioral data and machine learning algorithms of users. Taking Alibaba as an example, its intelligent recommendation system handles more than 1 billion recommendation requests per day, covering more than 90% of user interaction scenarios. The system deeply integrates explicit behavioral data such as user browsing records, search keywords, favorites, shopping carts, order history, etc., and at the same time combines implicit data such as device type, geographic location, and access time to build a dynamic user profile. For example, when a user searches for "outdoor sports equipment", the system not only recommends related products, but also predicts the user's interest in "carbon fiber hiking poles" based on his/her history of purchasing (e.g., preference for lightweight products) and real-time behavior (e.g., the type of products with the longest dwell time) through collaborative filtering algorithms and deep neural network models. Carbon fiber hiking poles" potential demand, and prioritize the display in the recommended list. In addition, the system also introduces a real-time feedback mechanism, if the user clicks on the recommended products and does not buy, the system will combine the evaluation data of the product, the price of competing products and the user's recent browsing alternatives, and dynamically adjust the recommendation strategy, such as pushing the "full-reduced coupon" or "combination set "To

improve the conversion rate. Alibaba's official data show that the intelligent recommendation system brings more than $100 billion GMV for the platform every year, in which the average conversion rate of recommended products reaches more than 10%, much higher than the non-recommended scenarios (Li, 2023).

4.2 User lifecycle management

Big data in e-commerce user lifecycle management, through the integration of the whole process of user behavioral data from the initial contact to the churn, to build a refined operation system, to achieve the maximization of user value mining. The core lies in the data-driven dynamic management of the five phases of the user life cycle (customer acquisition, activation, retention, revenue, and churn): in the customer acquisition phase, by analyzing the data of user source channels, search keywords and other data, identifying high-conversion-rate channels and optimizing the advertising strategy; in the activation phase, by using the user's behavioral data (such as frequency of use of the product and depth of interaction) after the first purchase, designing a personalized guidance In the retention stage, based on the user's historical purchase records, browsing preferences and other data, we build a user profile and implement a precise recommendation and membership system to enhance user stickiness; in the revenue stage, we analyze the user's purchase frequency, unit price, cross-buying behavior and other data to develop an exclusive marketing strategy for high-value users to enhance the value of the user's lifecycle; and in the churning stage, we predict the probability of user churn with the help of a machine learning model and combine it with user feedback to optimize the advertising strategy. In the churn stage, with the help of machine learning models, we predict the probability of user churn, locate the reasons for churn in combination with user feedback data, and take measures such as targeted offers and service improvements to win back users. Take Amazon Prime membership service as an example, it analyzes user purchase frequency, logistics preference, return rate and other data through big data, incorporates high-value users into the Prime system, and provides value-added services such as free delivery, video entertainment, etc., and dynamically adjusts membership rights and benefits according to changes in user behavior. Data shows that the average annual consumption of Prime members is more than two times that of non-members, and the churn rate is reduced by 40%.

4.3 Supply chain intelligent decision making

Big data in e-commerce supply chain intelligent decision making, through the integration of user behavior data, inventory status, logistics information and other multi-dimensional data, to build dynamic decision-making models, to achieve the whole chain of intelligence from demand forecasting to resource deployment. Take Amazon as an example, its supply chain intelligent decision-making system deeply integrates user behavior analysis and logistics network data: on the one hand, by analyzing user behavior data such as historical purchase records, search keywords, and product browsing time, combined with machine learning algorithms to predict regional demand trends for commodities, and deploy popular commodities to the front warehouse in advance (e.g., when predicting the surge in demand for outdoor products in a certain region, the system automatically increases the number of tents and hiking shoes in the local warehouse). (e.g., when the demand for outdoor products in a certain region is predicted to increase, the system automatically increases the inventory of tents and mountaineering shoes in the local warehouse); on the other hand, it uses IoT equipment to monitor the location of transport vehicles, the status of the goods, and external data such as the weather and traffic, so as to dynamically optimize the distribution path (e.g., if a route is congested due to a torrential downpour, the system immediately switches to the alternate route and adjusts the delivery time estimate). In addition, the system also analyzes the reasons for user returns, product evaluation and other feedback data, reverse drive supply chain upstream improvement (such as the discovery of a certain electronic product due to packaging fragility leads to a high return rate, the

system automatically triggers the production side of the packaging upgrade process). Data show that Amazon through the intelligent decision-making system to improve inventory turnover rate to the industry average of two times, at the same time, logistics costs accounted for the proportion of revenue from 8% to 6%, and its "predictive replenishment" model makes the popular goods out of stock rate dropped to less than 0.5% (Wang, 2023).

Conclusion

Big data has become the core driving force of e-commerce user behavior analysis, and its value runs through the whole chain of user interaction: through the integration of multi-source behavioral data, the intelligent recommendation system realizes the “1,000 people, 1,000 faces” precise reach, and the conversion rate is significantly improved; user life cycle management relies on the closed loop of data, accurately divides the user stage and customizes the operation strategy, and extends the user value cycle; supply chain management is also a key factor in the development of e-commerce. User life cycle management relies on the closed loop of data to accurately classify user phases and customize operation strategies to extend the value cycle of users; intelligent decision-making in the supply chain integrates demand prediction and logistics optimization to reduce costs and improve response speed. The upgrading of the whole chain from technology to organization further releases the potential of data and pushes e-commerce from traffic competition to data intelligence competition.

Conflict of Interest

The author declares that she has no conflicts of interest to this work.

References

Si, Y. (2025). Application of big data in e-commerce user behavior analysis. *Mall Modernization,* (08), 41–43.

Su, K., & Xu, D. (2025). Research on user behavior analysis and precision marketing of tourism e-commerce driven by big data. *Tourism Overview,* (05), 4–6.

Wei, J., & Wang, Y. (2019). Analysis and research on e-commerce user behavior based on big data of online shopping platform. *Journal of Hubei Institute of Technology, 35*(03), 34–38, 57.

Chen, S., & Ma, Y. (2024). E-commerce user behavior analysis and marketing strategy optimization. Mall Modernization, (16), 63–65.

Wu, S., & Zhou, M. (2025). Visual analysis of e-commerce user behavior. *Science and Industry, 25*(03), 335–341.

Ye, J. (2016). A personalized display method for commodities in Web environment [Monograph]. (Publisher not specified).

Li, F. (2023). Analysis of user behavior algorithms in e-commerce websites. *Electronic Technology, 52*(04), 202–203.

Wang, X. (2023). Detection of malicious payment users in e-commerce based on behavioral clustering. *Information and Computer (Theoretical Edition), 35*(03), 25–27.

How to Cite: Deng, Q. (2025). Application of Big Data in E-commerce User Behavior Analysis. *Journal of Global Humanities and Social Sciences*, *6*(5), 212-218
https://doi.org/ 10.61360/BoniGHSS252018640508