Application and Optimization of Cloud Computing in Big Data Processing

Zhong Chen*, Guoyan Yang and Feijiang Huang
School of Information and Communication Engineering,
Guangzhou Maritime University, Guangzhou, Guangdong, 510725, China

Keywords: Cloud Computing, Big Data Processing, E-Commerce Platforms, Application & Optimization.

Abstract: This paper studies the application of cloud computing in big data processing, taking the user behavior analysis of e-commerce platform as an example, aiming at the recommendation accuracy and user conversion rate of the platform. Based on the collection of user operation, product sales and regional data, the relationship between user behavior and conversion rate is analyzed. In this paper, it is necessary to build the platform architecture and framework, optimize the framework, and apply it, and finally conclude that the dwell time of purchase behavior is highly correlated with the conversion rate, and the high unit price products contribute significantly to sales. After analyzing the experimental data, it was concluded that the conversion rate was the highest in the northern region, while the conversion rate was lower in the southern region despite the high number of visits. Based on the comprehensive analysis, it was concluded that improving user dwell time and optimizing regional strategies could help improve its overall sales. This also shows that the application of cloud computing in big data processing is very effective, and effective improvements can be achieved through intelligent algorithm optimization.

1 INTRODUCTION

Because of the rapid development of e-commerce platforms, the analysis of user behavior data has become a key factor in improving their conversion rate (Ahmad, and Khan, et al. 2023). Many researchers have proposed that user conversion can be optimized based on recommendation algorithms, but the above methods do not perform well in processing big data, and cannot effectively deal with high-dimensional features and complex user behaviors. Some researchers have proposed a method based on improved recommendation algorithms, but the computing resources of the above methods are very efficient and cannot be applied on a large scale (Alshareef, 2023). This paper uses the combination of cloud computing technology and intelligent algorithms to analyze user behavior by using the powerful computing power of cloud computing (Arulmozhiselvan, and Uma, 2024). Through distributed processing, this paper hopes to improve the accuracy of platform recommendations. Another reason for choosing this method is that it can really process large amounts of data and is extremely good at predicting and optimizing user conversions in real time (He, Y. J and W. H. Ouyang, et al. 2023).

2 RELATED WORKS

2.1 Cloud Computing Technology

Cloud computing technology is a technical system that provides computing resources based on the network, which is characterized by on-demand allocation, scalability, and elastic computing. In a cloud computing environment, users do not need to manage physical hardware (Kang, and Deng, 2023), but can cope with computing needs in different periods as long as computing resources are dynamically adjusted. This type of on-demand allocation model is particularly suitable for large-scale data processing, such as real-time user data analysis on e-commerce platforms (Kumar, and Saini, 2024). Cloud computing is based on virtualization technology, which can process massive data requests

Chen, Z., Yang, G. and Huang, F.
Application and Optimization of Cloud Computing in Big Data Processing.
DOI: 10.5220/0013546200004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 1*, pages 485-491
ISBN: 978-989-758-763-4

at the same time, ensuring the efficient operation of the platform and improving resource utilization (Muhic, and Bengtsson, et al. 2023).

2.2 Application of Big Data Processing

The theory of big data processing mainly includes the storage, management, and analysis of massive data. The theory needs to be implemented through a distributed computing framework (Nagahawatta, and Warren, et al. 2024). For example, Spark can efficiently store and process data from different sources (Pericherla, 2023). Big data is characterized by large data volume, complex structure, and strong real-time performance, so it is necessary to improve processing efficiency based on parallel row computing and distributed storage. Big data processing also includes data mining techniques, such as extracting valuable information from user behavior to help platforms make personalized recommendations and market decisions (Pham, and Huynh-The, et al. 2023).

2.3 Mechanism of User Behavior Analysis

The theory of user behavior analysis is based on a detailed analysis of the various actions of users on the platform to understand the interests and preferences of users. Based on data mining and machine learning algorithms, the platform will be able to extract user behavior patterns from historical data, such as browsing and clicking, adding shopping carts, and purchasing behaviors. For example, user dwell time, product browsing order, click-through rate, bounce rate, etc. are all important parameters in behavioral analysis. Based on the analysis of the above parameters, e-commerce platforms can optimize the recommendation platform and improve the purchase conversion rate of its users, providing strong support for platform operations.

3 THE REALIZATION OF E-COMMERCE PLATFORMS COMBINING CLOUD COMPUTING AND BIG DATA PROCESSING

3.1 The Architecture of the E-Commerce User Behavior Analysis Platform

In the study, it is necessary to build a platform, because this article takes the analysis of user behavior of e-commerce platforms as an example, so it needs to have these blocks. It includes user interface blocks, request processing blocks, data storage blocks, data analysis blocks, recommendation engine blocks, and monitoring and feedback blocks. Among them, the user interface block is mainly responsible for interacting with the user. Based on the webpage, mobile terminal and other interfaces, it collects user input, such as search, click, purchase and other behaviors, and transmits user requests to the backend of the platform to ensure the friendliness and efficiency of the user experience. The request processing block is mainly responsible for receiving requests from the user interface block and scheduling appropriate services according to the type of request. Blocks are responsible for parsing user needs, such as product searches, recommendation requests, and interacting with back-end data processing blocks. The data storage block is responsible for storing all user data of the system, such as user profiles, browsing history, and purchase history. It is based on a distributed storage system to ensure the durability and fast reading of data

And it supports real-time query. The data analysis block is mainly responsible for real-time analysis of stored data and extracting useful user behavior patterns. Based on the analysis of users' browsing and purchase behavior, personalized recommendation results are generated, and data support is provided for subsequent user predictions. The recommendation engine block is mainly responsible for generating personalized product or content recommendations based on the historical behavior and current needs of its users. It leverages a well-established framework combined with real-time data to generate dynamic recommendations to improve user satisfaction. The monitoring and feedback block is mainly responsible for monitoring the operation status of the platform in real time, such as the response speed of user interaction and the accuracy of recommendation results. Based on the analysis of user feedback data, the block can dynamically adjust the platform to ensure the stability and optimization of the platform.

3.2 Construction of User Behavior Analysis Framework

Data sampling and distributed storage, in which based on data partitioning, cloud computing can distribute transaction data from users around the world to individual nodes for parallel computing, while reducing the pressure on a single node and improving processing efficiency. See Eq. (1) for details.

$$D_i = \{X_i, y_i\} \tag{1}$$

In the formula, D_i is data separation, which represents the partition of massive data by the e-commerce user behavior analysis platform, such as the user transaction records of the e-commerce website. Each partition is assigned to a different node for parallel processing, such as data retrieved from servers in different regions. The partitioned data is stored in parallel on platforms such as Hadoop. X_i Represents the feature set of the sample, which refers to the characteristics of user behavior on the e-commerce platform, such as the user's browsing time and the type of product clicked. The above characteristics will be used to predict the user's purchase behavior. For example, y_i indicates a purchase and 0 indicates that a user has not purchased. The above labels can guide the generation of decision trees when the framework is strengthened. Subsequently, the selection of random features and the optimization of cloud computing nodes were carried out. See (2).

$$F_{sub} = f_1, f_2, \ldots, f_k \tag{2}$$

In the formula, F is all available characteristics are represented, such as the user's click-through rate, product category, device type, etc. These characteristics help the framework understand user behavior and provide personalized recommendations. F_{sub} Represents a randomly selected set of features for each decision tree, such as browse time, commodity price, etc. Each tree is built using only a subset of these features, which increases the diversity of the framework. Based on the parallel processing of different feature sets at multiple nodes, the framework can strengthen multiple trees at the same time and shorten the strengthening time. In the cloud computing environment, random feature selection can help e-commerce platforms quickly process user behavior data, such as analyzing whether users will buy goods for certain characteristics, and improving the response speed and accuracy of their recommendation platforms.

However, the construction of decision trees and parallel processing are carried out. Specifically, in a distributed environment, e-commerce platforms can process multiple decision trees in parallel to capture different purchase patterns, such as different behaviors of users when browsing, adding to carts, and checking out. Based on distributed parallel computing, it can quickly process huge data sets and shorten prediction time. See Eq. (3) for this.

$$T_i = \mathit{Tree}(X_i, F_{sub}) \tag{3}$$

In the formula, T_i refers to the first decision tree, and each node processes specific data partitions and feature subsets in parallel to build a decision tree. In practice, each tree can capture different user behavior patterns, with some trees focusing on user browsing time and others focusing on price sensitivity. X_i Refers to user behavior data, such as features extracted from the order in which users visit the e-commerce platform. These features are used to enhance the decision tree and predict whether the user will buy the product or not.

3.3 Framework Enhancement and Optimization

In order to ensure the further optimization of the framework, it needs to be continuously strengthened until the iteration is completed and its performance in all aspects meets the requirements. The targeted reinforcement of the framework can improve the adaptability of the framework to different environments as much as possible, so as to better improve the practical application performance of the framework. In the process of reinforcement, it is necessary to optimize the number of trees and resource scheduling, and based on this, better reinforcement is carried out to meet the final requirements. See Eq. (4) for details.

$$N^* = \arg\min_N L(y, F(X; N)) \tag{4}$$

In the formula, N represents the number of decision trees in the framework. In the cloud computing environment, the number of adjustment trees can ensure the optimal performance of the framework, and at the same time, control the consumption of computing resources. In practice, increasing the number of trees can improve the accuracy of the framework, but it will consume more cloud computing resources, such as when processing user purchase predictions, if the number of trees is too small, certain behavioral patterns will be missed.

$L(y,\ F(X;\ N))$ Represents the loss function, which is used to measure the probability of user purchase and the error of actual purchase predicted by the framework. Based on the minimization loss function, the prediction accuracy of the framework can be improved. In this way, in the cloud computing environment, e-commerce platforms can.
Balance compute resources with the predictive accuracy of the framework and provide accurate purchase recommendations without sacrificing responsiveness.

In the optimization, it is necessary to complete the above content of tree depth optimization and complexity control, for which Eq. (5) can be seen.

$$d^* = \arg\min_d L(y, F(X; d)) \tag{5}$$

In the formula, d refers to the depth of the tree. In cloud computing big data processing, the depth of the tree controls the complexity of the framework. A tree that is too deep will bring an overfit, and a tree that is too shallow will not be able to capture enough features. By adjusting the depth of the tree, the framework will be able to capture key information and reduce its computational resource usage. Based on the depth of the optimization tree, the e-commerce user behavior analysis platform can effectively balance the complexity of the framework and computing resources, improve the accuracy of its prediction, and avoid the computational bottleneck when processing large-scale e-commerce data.

Frame compression and parameter optimization are also very important steps. In the optimization process, by performing this step, the e-commerce platform will effectively process the behavior data of hundreds of millions of users in the cloud computing environment, and significantly reduce storage and computing costs without affecting the prediction accuracy. See Eq. (6) for this.

$$M_{compressed} = \min_{T_i} \sum_{i=1}^{N} \| T_i \|_0 \tag{6}$$

In the formula, $M_{compressed}$ is the compressed frame. In the practical application of big data processing, the compression framework can significantly reduce the computing and storage overhead. For example, in the recommendation platform of an e-commerce platform, a compressed framework can reduce storage requirements and speed up response times, further improving the user experience. $\|T_i\|_0$ Indicates the number of parameters in the decision tree. Based on pruning or other compression techniques, T_i is the parameters of the decision tree will be effectively reduced, and the frame size will be reduced and the load of cloud computing will be reduced. In practice, this means that the framework can process user requests faster and improve the platform's real-time responsiveness.

3.4 Integration of E-Commerce User Behavior Analysis Platforms

The integration of this platform is mainly to define the block interface. First, the data interface and communication protocol between each block are determined to ensure the smooth transmission of information, such as the rapid interaction between the request processing block and the data storage block, and then effectively ensure the efficient acquisition and processing of data. Based on this, the block integration test is carried out, and all blocks are integrated into the test environment to verify the compatibility and data circulation of the interface. Based on simulated user requests, the stability of the platform can also be tested. In addition, in order to further optimize the performance of the platform, the performance of the platform should be optimized according to the test results after the integration test, specifically, it is necessary to improve the response speed of the data analysis algorithm and recommendation engine, and optimize the resource allocation to better reduce latency and improve the overall efficiency.

4 RESULTS AND DISCUSSION

4.1 The Case Results of Data Processing

The object of this application case is a large e-commerce enterprise M, whose main goal is to optimize its personalized recommendation platform and increase its overall sales based on the analysis of user behavior, product sales, and regional visit conversion rate. The user data of e-commerce enterprise M is large-scale, covering multiple regions and product categories, and at the same time, the user's operations on the enterprise platform have diversified characteristics, such as browsing, adding shopping carts, and finally purchasing. Based on data analysis, e-commerce company M hopes to improve the purchase conversion rate of users and make more accurate recommendations for different product categories. Using the e-commerce user behavior analysis platform built this time, e-commerce enterprise M will realize the application and optimization effect of cloud computing in big data processing on its own website. From 2020 to 2023, the e-commerce company M has developed extremely well, and its number of users continues to grow, as shown in Figure 1.

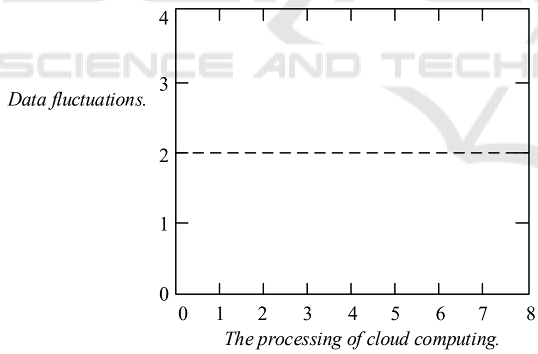

Figure 1: The development of the number of users of e-commerce company M from 2020 to 2023

4.2 Data Analysis Such as Basic User Operation Behavior

According to the case study, the dwell time of users is closely related to their actions. For example, based on user actions, the average dwell time of users who make a purchase is 10 minutes, while the dwell time of users who only click is 5 minutes. This indicates an increase in dwell time, representing a more likely user to complete a purchase, as shown in Table 1.

Table 1: User Operation Behavior and Dwell Time

The type of operation	Average dwell time (minutes)	Percentage of users (%)
Click Products	5	50
Add a shopping cart	7	30
Complete your purchase	10	20

Table 1 shows the analysis of users' operation behaviors on the platform, such as clicking, adding to shopping cart, and purchasing behaviors, the average dwell time of each operation, and the corresponding user proportion. Based on the table, it can be seen that user operation behavior and dwell time are the key contents of cloud computing in big data processing. Click on the product, add the shopping cart, and complete the purchase to promote the user's consumption, as shown in Figure 2.

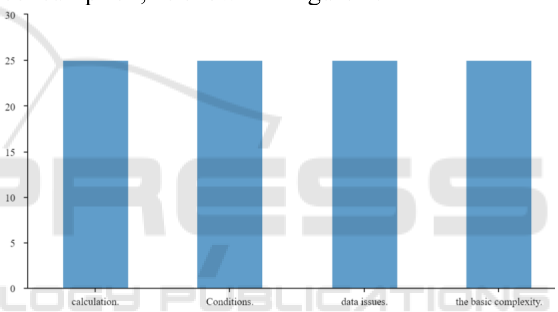

Figure 2: The promotion effect of clicking on the product, adding the shopping cart, and completing the purchase on the user's consumption

In this case, the sales of commodity categories can also have an impact on cloud computing big data processing. Specifically, electronics had the highest sales of 5,000 yuan, while books had the lowest average cart value of 35 yuan. This suggests that while books products have a lower per-transaction value, high-priced products contribute more to total sales. The specific sales of each product category and the average cart value are shown in Table 2.

Table 2: Sales of product categories

Product category	Total Sales (RMB)	Average cart value (RMB)
Electronics	5000	200
Books	1500	35
clothing	3000	75

Table 2 shows the sales data of product categories, such as the total sales of each type of product and the average value of the shopping cart, which provides a basis for optimizing product recommendations. Based on the above data, the e-commerce user behavior analysis platform will better understand the user's preference for different product categories, and provide support for subsequent product recommendations.

4.3 Visits and Conversion Rate Analysis by Region

From the analysis of visits and conversion rates in each region, the northern region has the highest conversion rate at 5.5%, although the southern region has a higher number of visits, but the conversion rate is lower. In addition, the southern region received 4,500 visits, but the conversion rate was the lowest at 3.2%, reflecting that users in the southern region are more willing to compare prices and consider other factors in the shopping process, as shown in Table 3.

Table 3: Visits and conversion rates by region

According to the type.	Visits	Conversion rate (%)
Optimize your phone.	3000	5.5
Actively optimize data.	4500	3.2
New phone.	4000	4.8

Table 2 shows the comparison of visits and conversion rates in different regions, with the aim of understanding user behavior patterns and purchasing decisions in each region to adjust regional marketing strategies. The relationship between visits and conversion rate and purchase decisions is shown in Figure 3.

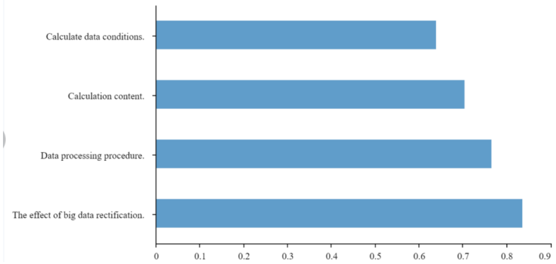

Figure 3: The relationship between a user's visits, conversion rate, and purchase decision

It can be seen that the conversion rate of users in the northern region is the strongest, and the conversion rate of users in the southern region is the weakest. This indicates that the purchase leads in the southern region need to be further optimized to improve their overall conversion rate. Based on the above data analysis, it can be concluded that user dwell time has a significant impact on purchase behavior, and the high unit price products in the product category contribute significantly to sales. In addition, regional marketing strategies need to be adjusted according to their conversion rates to improve overall sales performance. From the data, it can be fully proved that this study is valid.

5 CONCLUSIONS

Based on this paper, the advantages of cloud computing in big data processing have been fully verified, especially in the analysis of user behavior on e-commerce platforms. Cloud computing technology makes distributed storage and computing possible, effectively improving its data processing speed and resource utilization. Moreover, according to the research in this paper, based on the effective combination of parallel computing and intelligent algorithms, the platform can quickly analyze massive user behavior data and optimize the recommendation platform in real time. In the process of big data processing, the elastic scheduling of resources and the application of optimization algorithms will greatly improve the response speed and accuracy of the platform, and then provide high stability, high flexibility and effective technical support for the platform. In short, through the research of this paper, it can be found that the application based on cloud computing can not only help the platform achieve the efficiency of big data processing, but also bring higher sales performance and market competitiveness to the platform. The research in this paper contains a lot of data, but its cases are still limited, which makes it inevitably limited, and it can be further studied in the future to achieve effective optimization and scaling applications.

ACKNOWLEDGEMENTS

2022 Guangzhou Higher Education Teaching Quality and Teaching Reform Project - Department of Computer Basics (Project No.: 2022KCJYS017); 2022 Guangdong Higher Education Teaching Reform Project - Design and Implementation of the New Talent Training Program Management System Based on the OBE Concept (Project No.: C2206001163)

REFERENCES

Ahmad, A., R. A. Khan, S. U. Khan, H. S. Alwageed, A. A. Al-Atawi, and Y. M. Lee. (2023). Green cloud computing adoption challenges and practices. a client's perspective-based empirical investigation. Cognition Technology & Work, 25(4), 427-446.

Alshareef, H. N. (2023). Current Development, Challenges and Future Trends in Cloud Computing. A Survey. International Journal of Advanced Computer Science and Applications, 14(3), 329-338.

Arulmozhiselvan, L., and E. Uma. (2024). Cloud and Fog Computing Amalgamation for Data Agitation and Guard Intensification in Health Care Applications. Ksii Transactions on Internet and Information Systems, 18(3), 685-703.

He, Y. J., W. H. Ouyang, S. L. Li, L. Wang, J. Zhou, W. W. Su, S. Z. Li, D. H. Mei, Y. Shi, Y. X. Jin, C. L. Li, and Y. H. Ren. (2023). Cloud computing data privacy protection method based on blockchain. International Journal of Grid and Utility Computing, 14(5), 480-492.

Kang, H. Y., and J. Deng. (2023). A cross encryption scheme for data security storage in cloud computing environment. International Journal of Internet Protocol Technology, 16(1), 1-10.

Kumar, K., and D. K. Saini. (2024). Security issues in cloud, fog and edge computing models and suggested solutions. Journal of Discrete Mathematical Sciences & Cryptography, 27(4), 1247-1257.

Muhic, M., L. Bengtsson, and J. Holmström. (2023). Barriers to continuance use of cloud computing. Evidence from two case studies. Information & Management, 60(5).

Nagahawatta, R., M. Warren, S. Salzman, and S. Lokuge. (2024). Towards an Understanding of Cloud Computing Adoption in SMEs. The Role of Security and Privacy Factors. International Journal of Cyber Warfare and Terrorism, 14(1).

Pericherla, S. S. (2023). Cloud Computing Threats, Vulnerabilities and Countermeasures. A State-of-the-Art. Isecure-Isc International Journal of Information Security, 15(1), 1-58.

Pham, X. Q., T. Huynh-The, D. S. Kim, T. D. Nguyen, and E. N. Huh. (2023). Distributed Cloud Computing. Architecture, Enabling Technologies, and Open Challenges. Ieee Consumer Electronics Magazine, 12(3), 98-106.