Scholar Publishing Group
International Journal of Big Data Intelligent Technology
https://doi.org/10.38007/IJBDIT.2026.070102
ISSN 2790-0932 Vol. 7, Issue 1: 9-17

Performance Optimization and Improvement of Advertising Machine Learning Platform Based on Distributed Systems

Yixian Jiang

Carnegie Mellon University, Information Networking Institute, Pittsburgh, PA, 15213, USA

Keywords: Distributed system; Advertising machine learning; Performance optimization; Data preprocessing; Distributed training

Abstract: With the rapid advancement of advertising technology, advertising intelligent algorithm platforms are facing increasing computing and storage challenges, especially in the context of distributed architecture. Improving platform performance has become a key and urgent task. This study is based on the basic framework of distributed systems, and deeply explores the performance constraints of advertising intelligent algorithm platforms in areas such as data storage, computing resource allocation, algorithm training, and network broadband. Targeted optimization measures are proposed. Specific strategies include adopting advanced data processing architectures, implementing flexible resource allocation and management, using distributed algorithm training architectures to accelerate model training processes, and optimizing data transmission processes to enhance platform performance. After experimental testing, the optimized platform exhibits better efficiency and lower latency in processing large amounts of data and high-intensity computing tasks, comprehensively improving the performance level of advertising intelligence algorithms. This study provides an effective solution for improving the performance of advertising intelligent algorithm platforms and has high practical value.

1. Introduction

When dealing with the massive user data parsing, real-time decision-making, and customized recommendations of advertising machine learning platforms, the strong demand for computing power and storage resources has become particularly prominent. Distributed architecture, as a key solution for solving large-scale data processing and high concurrency computing tasks, has been widely deployed in the field of advertising machine learning. In a distributed environment, advertising machine learning platforms still face many challenges such as data storage limitations, computational performance bottlenecks, uneven resource distribution, and long model training times, all of which affect the overall performance of the system. Exploring and improving the performance of advertising machine learning platforms to address these challenges has become a focus of academic attention. This study aims to analyze these difficulties and propose specific

performance improvement measures to enhance the efficiency and scalability of advertising machine learning platforms.

2. Characteristics of Distributed Systems

Distributed systems, with their unique characteristics, as shown in Figure 1, exhibit unique advantages in handling complex tasks and large-scale datasets. Its parallel processing capability is particularly outstanding, as multiple nodes can synchronously execute different tasks, improving computational efficiency. For advertising machine learning platforms, they must handle massive amounts of user data and advertising content, while distributed systems reduce processing time by distributing computing tasks to numerous nodes for parallel operations. Scalability is also a key attribute of distributed systems. Faced with the continuous increase in data scale and processing requirements, the system can flexibly expand computing resources. By adding more nodes, the system can handle a larger amount of tasks and data without interfering with existing services. Distributed systems have shown strong adaptability and advantages in the application of advertising machine learning platforms, effectively solving performance bottlenecks and ensuring service efficiency and stability.

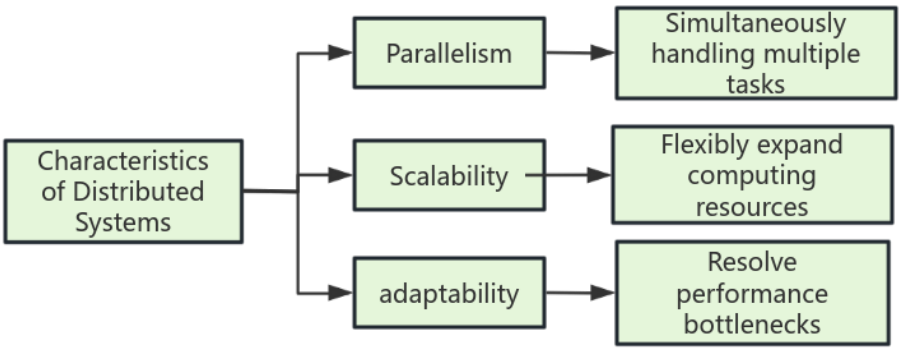

Figure 1. Characteristics of Distributed Systems

3. Performance issues of advertising machine learning platforms

3.1 Bottlenecks in large and complex data storage and computation

In advertising machine learning platforms, massive amounts of advertising information and related data need to be efficiently processed and saved, including key data such as user behavior tracking, ad click through rates, and conversion tracking. These data volumes are enormous, diverse in types and structures, covering everything from textual information to images, videos, and data captured by various sensors. Faced with vast and complex data sets, traditional single machine processing and storage solutions are no longer able to efficiently meet the requirements, and there are obvious performance bottlenecks. Especially in terms of data storage, with the continuous expansion of user scale, storage demand is exponentially increasing, which puts enormous pressure on databases and file systems, limits storage space, and affects data read and write efficiency. Traditional relational databases are difficult to handle complex data structures, while distributed file systems and NoSQL databases, although having some scalability, still face challenges in data consistency and transaction processing, which increases the overall complexity of the system.

3.2 Unequal allocation of computing resources and excessive system load

In advertising intelligent learning systems, the imbalance of resource allocation is one of the key obstacles that constrain the efficiency of system operation. Given that advertising systems need to cope with massive data streams and complex intelligent algorithm models, resource allocation often deviates, putting immense pressure on certain computing nodes and affecting the overall performance of the system. The load faced by advertising intelligent learning systems is composed of a mixture of numerous tasks and requests, each with different computational requirements and data processing volumes. If the resource scheduling strategy of the system cannot monitor and adjust the load status of each node in a timely manner, some resource intensive tasks may excessively consume computing resources, compress the computing space of other tasks, and prolong the entire computing process. During periods of surge in traffic, uneven allocation of resources is particularly prominent, which may lead to a decrease in system efficiency and, in severe cases, may cause system paralysis or temporary service interruption. If the allocation mechanism of computing resources fails to reasonably consider the priority differences between tasks, the required computing load, and the specific requirements for data access, it may result in some tasks being unable to obtain necessary computing resources, affecting the processing capacity and response time of the entire system.

3.3 Model training time is too long

On advertising intelligence algorithm platforms, model training is a resource intensive activity, especially when dealing with massive datasets, where the training cycle is often extremely long. Given the massive amount of advertising data, such as user base, behavior history, and real-time click streams, relying solely on traditional single machine training methods is no longer sufficient to meet the requirements of fast and efficient processing. Even with the application of distributed computing architecture, excessively long model training time is still an urgent problem that the platform needs to solve. These advertising algorithm models are designed with complexity, involving numerous feature processing, parameter optimization, and model validation steps. Faced with a huge amount of data, model training must go through multiple iterations, adjust various hyperparameters, and perform cross validation. These operations are not only time-consuming, but also require extremely high computing power. Even with concurrent computing using multiple nodes, time constraints are difficult to completely overcome. Especially when applying deep learning techniques, the increase in training time is almost exponential, which undoubtedly poses a serious challenge for advertising systems that pursue instant feedback. The excessively long model training cycle affects the response speed and work efficiency of the advertising intelligent platform, and may also lead to resource waste, reduce the system's response sensitivity, and affect the effectiveness of advertising placement and user interaction experience.

3.4 Network bandwidth limitations in distributed computing

In distributed advertising intelligent algorithm systems, network bandwidth plays a key role in constraining overall performance due to its highly dependent architecture on large-scale data exchange. Faced with massive advertising data processing and intelligent algorithm computing tasks, frequent data exchange between nodes and broadband limitations may evolve into performance bottlenecks in the system, thereby reducing computing speed. In distributed processing architecture, the demand for data transmission is extremely large, especially during the execution of large-scale data caching, algorithm training, and parameter update operations. In advertising intelligent algorithm systems, achieving real-time data transfer between computing units is

particularly crucial, especially in distributed training mode where parameter synchronization and gradient information exchange between nodes will generate massive data traffic. Once the network broadband is insufficient, the data transmission rate will be limited, resulting in communication delays among nodes and affecting the overall computational efficiency of the system. The distributed network broadband limitation not only slows down data transmission speed, but may also cause a backlog of computing tasks, weaken the overall performance of the system, and become a major obstacle to improving the efficiency of advertising intelligent algorithm systems.

4. Performance optimization and improvement strategies for advertising machine learning platforms

4.1 Introducing an efficient data preprocessing framework to reduce data processing time

Preprocessing data in advertising machine learning platforms is the core step in optimizing model training effectiveness and speed. This process generally involves data purification, attribute filtering, data standardization, and filling in data gaps, which often require a significant amount of computing power and time cost. Especially when dealing with the large, ever-changing, and continuously updated user data in advertising systems, traditional data processing methods cannot meet the requirements of efficient and real-time processing. Adopting advanced data processing architecture can shorten the time required for data processing and improve the overall operational efficiency of the platform.

The advanced data preparation workflow supports distributed job execution of data, by dispersing the relevant operations of data preparation across numerous processing units for synchronous execution. This strategy improves the speed of data processing and alleviates the processing latency caused by the performance limitations of a single processing unit. Data preparation systems generally have scalable features that can automatically allocate computing power according to the growth of data size, ensuring that system performance does not decrease due to the expansion of data volume. The current data preparation system integrates various efficient algorithms and optimization methods, including data deduplication, compression, and feature extraction, which improves the efficiency of the data preparation stage and saves computing resources. The normalization process in data preprocessing can also be optimized through efficient algorithms. For example, the commonly used data normalization method in standardization is:

$$X' = \frac{X - \mu}{\sigma} \tag{1}$$

In formula (1), X is the original data, μ is the mean of the data, σ is the standard deviation of the data, and X'is the normalized data. By adopting this advanced data preprocessing architecture, advertising machine learning platforms can achieve tasks such as data cleaning, format conversion, and standardization in a relatively short period of time, improving the efficiency of data processing. This provides stronger data support for the subsequent model construction and real-time advertising evaluation. This strategy shortens the data processing cycle, reduces the burden on system resources, and further enhances the overall performance and responsiveness of the system.

4.2 Dynamic Resource Scheduling and Management

In the intelligent advertising processing system, flexible allocation and regulation of resources is one of the key mechanisms under distributed architecture. It can track the running load of each computing node in real time, optimize the configuration of computing power, and ensure the smooth operation of the system. The core of this resource management strategy lies in automatically

adjusting resource allocation based on real-time workloads and system conditions. This scheduling mechanism improves the effective utilization of computing resources and can flexibly adjust resource allocation plans based on task importance, execution time, and computing requirements. In distributed intelligent advertising systems, given the unpredictability of advertising information and computing tasks, dynamic resource allocation can flexibly adjust resources based on real-time workload and task characteristics, ensuring performance stability even in high concurrency environments.

Taking the algorithm training work involved in intelligent advertising placement as an example, an infrastructure with numerous computing nodes has different processing capabilities, and the training tasks assigned to each model have differences in computation and resource consumption. With the help of a flexible task allocation mechanism, the system can decide whether to assign tasks to the most suitable nodes or temporarily allocate additional computing power based on the urgency of the tasks, current computational pressure, and the operational status of the nodes. For example, when the computational requirements of a model suddenly increase, the scheduling system will automatically expand resources to speed up processing. When the pressure of the task decreases, the system will automatically reduce the allocated resources and minimize unnecessary resource consumption. The following table shows the changes in computational resource utilization efficiency and task completion time of advertising machine learning platforms before and after the introduction of dynamic resource scheduling:

Table 1 Resource Utilization Efficiency of Advertising Machine Learning Platform

Scheduling Strategy	Total number of tasks	Total number of nodes	Average load balancing degree	Total calculation time (hours)	Average utilization rate of computing resources (%)	Task completion time (minutes)
Static resource allocation	1000	10	65%	150	70%	45
Dynamic resource scheduling optimization	1000	10	95%	120	85%	35

Observing the data table, it can be found that after adopting the dynamic resource scheduling optimization strategy, the resource utilization efficiency of the system increased, and the workload balancing of nodes improved to 95%, which is a qualitative leap compared to the traditional static resource allocation of 65%. Thanks to the implementation of dynamic resource scheduling and management, advertising machine learning platforms are able to flexibly adjust resource allocation based on real-time computing task requirements, solving the problems of resource surplus and uneven load, accelerating system response speed and improving processing efficiency, providing stronger technical support for advertising analysis and customized push.

4.3 Accelerating Model Training through Distributed Training Framework

On the machine learning platform of intelligent advertising systems, model training is a task that requires extremely high computing power, especially when dealing with huge amounts of data. The traditional way of training relying on a single computer often fails to meet efficiency and time

standards. In order to accelerate the development of models, distributed training architectures have been widely applied in the field of intelligent advertising. This architecture disperses training tasks across numerous computing units, unleashing the powerful power of parallel processing and accelerating model generation efficiency. This strategy improves computational efficiency and can handle large datasets, effectively shortening the training cycle of the model.

The advertising platform uses deep learning models for advertising recommendation tasks, which include multiple levels of neural networks. In a distributed training framework, training data is divided into several batches, and each batch is allocated to different nodes for parallel processing. The gradient calculated for each node is $\nabla J(\theta_i)$, where $\nabla J(\theta_i)$ is the gradient of the i node. Through data parallelism, the gradient calculation results of each node will be summarized on a parameter server and then globally updated:

$$\theta^{t+1} = \theta^t - \eta \cdot \frac{1}{N} \sum_{i=1}^{N} \nabla J(\theta_i) \tag{2}$$

In formula (2), θt is the model parameter for the current iteration, η is the learning rate, N is the number of nodes, $\nabla J(\theta_i)$ is the gradient of the i node, and θ^{t+1} is the global parameter for the next iteration. With the help of distributed gradient descent strategy, the data processing and computational communication ability of the advertising intelligent learning system is enhanced during the training phase, improving the efficiency of training. By utilizing a distributed training architecture, the system is able to parallelize massive amounts of data, shorten model training time, and enhance overall system performance. This optimization method is particularly suitable for handling advertising push tasks containing large amounts of data, high-dimensional feature vectors, and complex algorithms, improving the training efficiency and prediction accuracy of the model.

4.4 Compressed transmission of data using efficient transmission protocols

On the algorithm platform of intelligent advertising, the use of advanced transmission mechanisms and data reduction techniques can improve the speed of data transmission, shorten the delay in the transmission process, reduce the burden on the system, and enhance the overall performance of the platform. This efficient transmission mechanism not only ensures the stability and smoothness of data transmission, but also improves transmission efficiency by improving the composition of data packets and reducing unnecessary data transmission. For example, transport protocols such as gRPC and Apache Kafka have been widely used in distributed architectures, supporting efficient end-to-end communication and enabling concurrent processing and asynchronous transmission of data streams. Combined with data compression methods such as LZ4, Snappy, etc., compressing the data before sending greatly reduces the demand for network bandwidth and accelerates data transmission speed. The following table shows the changes in data transmission efficiency of advertising machine learning platforms before and after adopting efficient transmission protocols:

Table 2 Changes before and after efficient transmission protocol

Transmission Protocol	Original data size (MB)	Compressed data size (MB)	Transmission time (seconds)	Network bandwidth utilization rate (%)
TCP/IP	100	100	60	75%
gRPC + Snappy	100	30	25	95%

From Table 2, it can be seen that the use of efficient transmission protocols combined with data compression has improved transmission efficiency, reduced the consumption of network resources,

and enhanced the system's response speed and processing capabilities. By reducing the data transmission volume and utilizing efficient communication protocols, the advertising intelligent analysis system reduces data transmission time and bandwidth usage, enhancing system performance. This improves the system's operational efficiency and user experience in the face of high concurrency scenarios, processing large-scale datasets, and advertising push and evaluation work that requires quick response.

5. Conclusion

This article focuses on distributed architecture and explores in depth the performance limitations of machine learning platforms in the advertising field, with a particular emphasis on the challenges faced in key areas such as data processing, resource allocation, model iteration, and network communication. A series of targeted improvement measures have been proposed, including adopting advanced data preprocessing architecture, improving resource management methods, implementing distributed acceleration model learning, and optimizing data transmission processes. The effective implementation of these measures enhances the performance of advertising machine learning platforms in handling massive amounts of data and carrying high load tasks, bringing stronger technical support to the advertising field. In the future, with the continuous development of hardware devices and algorithm technology, the performance of distributed advertising machine learning platforms is expected to achieve greater leaps, laying a more solid foundation for the future development of advertising technology.

Reference

[1] *El K N, Belangour A. Research Intelligent Precision Marketing of Insurance Based on Explainable Machine Learning: A Case Study Of An Insurance Company. journal of theoretical and applied information technology, 2024, 102(6):2598-2607.*

[2] *Beauvisage T, Beuscart J S, Coavoux S, et al. How online advertising targets consumers: The uses of categories and algorithmic tools by audience planners:. New Media & Society, 2024, 26(10):6098-6119.*

[3] *Aljabri M, Mohammad R M A. Click fraud detection for online advertising using machine learning. Egyptian Informatics Journal, 2023, 24(2):341-350.*

[4] *Zhou F, Jiang Y, Chai L Y. Product consumptions meet reviews: Inferring consumer preferences by an explainable machine learning approach. Decision Support Systems, 2024, 177(Feb.):114088. 1-114088. 15.*

[5] *Su C, Wei J, Lei Y, et al. Empowering precise advertising with Fed-GANCC: A novel federated learning approach leveraging Generative Adversarial Networks and group clustering. PLoS ONE, 2024, 19(4).*

[6] *Xiu, L. (2025, June). Research on Personalized Recommendation Algorithms in Modern Distance Education Systems. In 2025 IEEE 3rd International Conference on Image Processing and Computer Applications (ICIPCA) (pp. 2019-2024). IEEE.*

[7] *Xu, D. (2025). Integration and Optimization Strategy of Spatial Video Technology in Virtual Reality Platform. International Journal of Engineering Advances, 2(3), 131-137.*

[8] *Huang, J. (2025). Adaptive Reuse of Urban Public Space and Optimization of Urban Living Environment. International Journal of Engineering Advances, 2(4), 9-17.*

[9] *Zhou, Y. (2025). Using Big Data Analysis to Optimize the Financing Structure and Capital Allocation of Energy Enterprises. Economics and Management Innovation, 2(7), 8-15.*

[10] Zhang, Q. (2025). Use Computer Vision and Natural Language Processing to Optimize Advertising and User Behavior Analysis. Artificial Intelligence and Digital Technology, 2(1), 148-155.

[11] Wang, Y. (2025). Research on Early Identification and Intervention Techniques for Neuromuscular Function Degeneration. Artificial Intelligence and Digital Technology, 2(1), 163-170.

[12] Wu, H. (2025). The Challenges and Opportunities of Leading an AI ML Team in a Startup. European Journal of AI, Computing & Informatics, 1(4), 66-73.

[13] Ren, B. (2025). Cross Modal Data Understanding Based on Visual Language Model. European Journal of AI, Computing & Informatics, 1(4), 81-88.

[14] Shen, D. (2025). Innovative Application of AI in Medical Decision Support System and Implementation of Precision Medicine. European Journal of AI, Computing & Informatics, 1(4), 59-65.

[15] Liu, X. (2025). Use Generative Al and Natural Language Processing to Improve User Interaction Design. European Journal of AI, Computing & Informatics, 1(4), 74-80.

[16] Liu, F. (2025). Localization Market Expansion Strategies and Practices for Global E-commerce Platforms. Strategic Management Insights, 2(1), 146-154.

[17] Hu, Q. (2025). Research on the Combination of Intelligent Management of Tax Data and Anti-Fraud Technology. Strategic Management Insights, 2(1), 139-145.

[18] Hua, X. (2025). Key Indicators and Data-Driven Analysis Methods for Game Performance Optimization. European Journal of Engineering and Technologies, 1(2), 57-64.

[19] Ren, B. (2025). Deep Learning and Anomaly Detection in Predictive Maintenance Platform. European Journal of Engineering and Technologies, 1(2), 65-71.

[20] Hui, X. (2025). Research on the Application of Integrating Medical Data Intelligence and Machine Learning Algorithms in Cancer Diagnosis. International Journal of Engineering Advances, 2(3), 101-108.

[21] Hui, X. (2025). Utilize the Database Architecture to Enhance the Performance and Efficiency of Large-Scale Medical Data Processing. Artificial Intelligence and Digital Technology, 2(1), 156-162.

[22] Jingzhi Yin. Research on Financial Time Series Prediction Model Based on Multi Attention Mechanism and Emotional Feature Fusion. Socio-Economic Statistics Research (2025), Vol. 6, Issue 2: 161-169

[23] Dingyuan Liu. Measuring the Sensitivity of Local Skill Structures to AI Substitution Risks Based on Occupational Task Decomposition. Socio-Economic Statistics Research (2025), Vol. 6, Issue 2: 177-184

[24] Yiting Hong. An Efficient Federated Graph Neural Network Framework for Cross-Enterprise Business Analysis. Socio-Economic Statistics Research (2025), Vol. 6, Issue 2: 170-176.

[25] Jiahe Sun. Research on Financial Systemic Risk Measurement Based on Investor Sentiment and Network Text Mining. Socio-Economic Statistics Research (2025), Vol. 6, Issue 2: 185-193.

[26] Thanh-Huyen Truong. Research on the Mechanism of E-commerce Model Innovation Driven by Digital Technology. International Journal of Big Data Intelligent Technology (2025), Vol. 6, Issue 2: 171-178

[27] Chen, X. (2025). Research on AI-Based Multilingual Natural Language Processing Technology and Intelligent Voice Interaction System. European Journal of AI, Computing & Informatics, 1(3), 47-53.

[28] Qi, Y. (2025). Data Consistency and Performance Scalability Design in High-Concurrency Payment Systems. European Journal of AI, Computing & Informatics, 1(3), 39-46.

[29] Fu, Y. (2025). The Push of Financial Technology Innovation on Derivatives Trading Strategy Optimization. European Journal of Business, Economics & Management, 1(4), 114-121.

[30] Li, J. (2025). High-Performance Cloud-Based System Design and Performance Optimization Based on Microservice Architecture. European Journal of AI, Computing & Informatics, 1(3), 77-84.