Optimizing Resource Allocation using Federated Learning with Edge Computing

Chaohui Zhu

School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, China
zhuchh9@mail2.sysu.edu.cn

Abstract. Resource allocation in parallel and distributed systems is a critical challenge, particularly in environments where data privacy, latency, and bandwidth are of concern. This paper proposes a novel approach to resource allocation by integrating Federated Learning (FL) with Edge Computing. The proposed FL-based model optimizes resource distribution by enabling decentralized learning across edge devices without requiring centralized data aggregation. We evaluate the performance of our approach through simulations, demonstrating significant improvements in latency, throughput, and energy efficiency compared to traditional methods.

Keywords: Federated Learning · Edge Computing.

1 Introduction

Parallel and distributed systems have become the backbone of modern computing, enabling large-scale processing and efficient utilization of resources. However, the growing complexity of these systems presents significant challenges in resource allocation, particularly when considering constraints such as data privacy, latency, and bandwidth. Traditional resource allocation methods often rely on centralized control, which can lead to bottlenecks and inefficiencies.

Federated Learning (FL) offers a promising solution by allowing distributed devices to collaboratively learn a shared model without exchanging raw data. When combined with Edge Computing, which brings computation closer to the data source, FL can optimize resource allocation in parallel and distributed systems. This paper explores the integration of FL with Edge Computing to develop a resource allocation model that enhances performance while addressing privacy and bandwidth concerns.

2 Model

The proposed model integrates Federated Learning with Edge Computing to optimize resource allocation in parallel and distributed systems. The architecture consists of the following components:

- **Edge Devices:** These devices perform local model training using their own data. The trained models are then sent to a central coordinator for aggregation.
- **Central Coordinator:** This entity aggregates the locally trained models from the edge devices to create a global model. The global model is then distributed back to the edge devices for further training.
- **Resource Manager:** The Resource Manager dynamically allocates computational resources based on the current workload and network conditions, ensuring optimal performance.

The FL process is iterative, with each round involving local training at the edge devices, model aggregation at the central coordinator, and resource reallocation by the Resource Manager. This iterative process continues until the global model converges or a predefined number of iterations is reached.

Algorithm 1 FL-based Resource Allocation in Parallel and Distributed Systems

Initialization:
Distribute an initial global model $M_{\text{global}}^{(0)}$ to all edge devices.
for each round $t = 1, 2, \ldots, T$ **do**
    Local Training (at each Edge Device i):
    for each edge device i in parallel **do**
        Use local data D_i to train the model $M_i^{(t)}$.
        Compute the model update $\Delta M_i^{(t)} = M_i^{(t)} - M_{\text{global}}^{(t-1)}$.
    end for
    Model Aggregation (at Central Coordinator):
    Collect model updates $\Delta M_i^{(t)}$ from all edge devices.
    Distribute the updated global model $M_{\text{global}}^{(t)}$ to all edge devices.
    Resource Reallocation (by Resource Manager):
    Evaluate the current workload $W_i^{(t)}$ and network conditions at each edge device i.
    Adjust resource allocation $R_i^{(t)}$ across edge devices to optimize performance
    Convergence Check:
    if the global model $M_{\text{global}}^{(t)}$ has converged or $t = T$ **then**
        Break loop.
    else
        Proceed to the next round.
    end if
end for
Final Output:
Deploy the converged global model $M_{\text{global}}^{(T)}$ across the edge devices.
Optimize resource allocation for ongoing operations.

3 Algorithm

The proposed algorithm for Federated Learning (FL)-based resource allocation in parallel and distributed systems is detailed below. The algorithm integrates FL with Edge Computing to optimize resource distribution across the system.