

A Fine-Grained Scheduling Algorithm for Heterogeneous Supercomputing Clusters Based on Graph Convolutional Networks and Proximal Policy Optimization

Authors : Jiahao Zhou, Lei Wang

Abstract : In heterogeneous supercomputing clusters, designing an efficient scheduling strategy is crucial for enhancing both energy efficiency and workflow execution performance. The dynamic allocation and reclamation of computing resources are essential for improving resource utilization. However, existing studies often allocate fixed resources to jobs prior to execution, maintaining these resources until job completion, which overlooks the importance of dynamic scheduling. This paper proposes a heterogeneous hierarchical fine-grained scheduling algorithm (HeHiFiS) based on graph convolutional networks (GCN) and proximal policy optimization (PPO) to address issues such as prolonged workflow completion times and low resource utilization in heterogeneous supercomputing clusters. Specifically, GCN is employed to extract task dependency features as part of the state information, and the PPO reinforcement learning algorithm is then used to train the scheduling policy. The trained scheduling policy dynamically adjusts scheduling actions during operation based on the continuously changing states of tasks and computing resources. Additionally, we developed a heterogeneous scheduling simulation platform to validate the effectiveness of the proposed algorithm. Experimental results indicate that HeHiFiS, by incorporating resource inheritance and intra-task parallel mechanisms, significantly improves resource utilization. Compared to existing scheduling algorithms, HeHiFiS achieves over a 50% improvement in both job completion and response performance metrics.

Keywords : heterogeneous, dynamic scheduling, GCN, PPO

Conference Title : ICSLP 2025 : International Conference on Speech and Language Processing

Conference Location : Algiers, Algeria

Conference Dates : March 24-25, 2025