面向大模型训练的算力资源调度策略研究

王冀远
杭州蚂蚁酷爱科技有限公司　浙江　杭州　310012

摘　要：面向大模型训练的算力资源调度策略研究聚焦于应对大模型训练中算力需求激增、资源动态变化及异构适配等挑战。研究提出分层调度模型，结合动态优先级分配、异构资源适配及通信-计算重叠优化策略，实现资源高效利用。同时，引入能耗感知调度与强化学习驱动的自适应调度，平衡能耗与效率，提升调度智能化水平，为大模型训练提供可持续优化的算力支持。

关键词：大模型训练；算力资源；调度策略

引言：在人工智能迅猛发展的当下，大模型凭借强大的性能在诸多领域广泛应用，但其训练过程对算力需求呈指数级增长。传统算力调度策略在应对大模型训练时，暴露出忽略动态特性、异构适配性差、能耗效率失衡等问题。如何设计高效的算力资源调度策略，以充分挖掘集群算力、降低能耗、提升训练效率，成为推动大模型发展的关键，本研究正是基于此展开深入探讨。

1 相关工作综述

1.1 大模型训练技术现状

（1）分布式训练架构：为突破单设备算力瓶颈，分布式训练成为主流方案，主要分为三类。数据并行将训练数据拆分至多个设备，各设备独立计算梯度后同步，适配大规模数据场景；模型并行将大模型参数拆分部署，解决单设备内存不足问题；流水线并行按训练步骤拆分任务，实现设备间并行执行，提升训练吞吐量。（2）通信与计算协同优化：通信延迟是分布式训练的核心瓶颈，相关优化技术持续突破。AllReduce算法实现多设备梯度高效聚合，降低通信开销；梯度压缩技术通过量化、稀疏化等方式减少通信数据量，在保证训练精度的前提下，提升通信与计算协同效率[1]。

1.2 传统算力调度策略

（1）基于静态分配的调度：采用固定分配规则，无需实时感知负载状态。FIFO按任务提交顺序调度，实现简单但易导致算力浪费；RoundRobin按固定顺序循环分配算力，保证任务公平性，但未考虑任务算力需求差异。（2）基于负载均衡的调度：以均衡设备负载为核心目标。Min-Min优先调度执行时间最短的任务，Max-Min优先调度执行时间最长的任务，均能缓解负载不均问题，但未结合训练任务的动态特性设计。

1.3 现有问题与不足

（1）忽略大模型训练的动态特性：大模型训练过程中，梯度同步、Checkpoint存储等操作会产生动态开销，现有策略多按静态场景设计，无法动态适配开销变化，导致调度效率下降。（2）未考虑异构算力的适配性：当前训练多采用GPU/TPU/CPU混合集群，不同设备算力、通信特性差异较大，现有调度策略未针对性适配设备异构性，难以充分发挥集群整体算力。（3）能耗与效率的权衡不足：大模型训练能耗极高，现有策略多优先追求训练效率，未合理平衡能耗与效率，导致算力利用效率低、能耗浪费严重，不符合绿色计算需求。

2 大模型训练的算力需求分析

2.1 计算任务特征

（1）计算密集型：大模型训练的核心是海量矩阵运算与反向传播过程，涉及数十亿甚至上万亿参数的迭代更新，计算量呈指数级增长。其中，前向传播中的矩阵乘法、激活函数运算，以及反向传播中的梯度求解、参数更新，均需要占用大量计算资源，对设备的浮点运算能力提出极高要求，GPU、TPU等专用计算设备成为主流选择，其计算性能直接决定了训练效率。（2）通信密集型：大规模分布式训练中，多设备、多节点间的参数同步、数据传输的通信开销日益突出，成为制约训练速度的关键因素。训练过程中，各节点需实时同步梯度参数、共享训练数据，尤其是在数据并行、模型并行架构下，参数同步频率高、数据量大，对网络通信的带宽和延迟提出严苛要求，通信效率直接影响整体训练的吞吐量[2]。

2.2 资源瓶颈识别

（1）计算资源：计算资源瓶颈主要体现在GPU利用率不均衡与内存带宽不足两方面。一方面，大模型训练中，部分任务存在计算负载不均现象，导致部分GPU处于空闲或低负载状态，资源利用率偏低；另一方面，海量参数的存储与读取需要高速内存带宽支撑，内存带宽不足会导致数据读写延迟增加，出现“计算等待数据”

的瓶颈，拖累整体计算效率。（2）通信资源：网络带宽有限与通信延迟过高是通信资源的核心瓶颈。分布式训练中，多节点间的梯度同步、数据传输需要占用大量网络带宽，当带宽不足时，会出现数据传输阻塞；同时，节点间的通信延迟会累积叠加，尤其是在跨节点、跨集群训练场景下，延迟问题更为突出，严重影响参数同步效率，降低训练吞吐量。

2.3 动态性来源

（1）训练阶段差异：大模型训练需经历预热、收敛、微调三个核心阶段，各阶段的算力需求差异显著。预热阶段主要进行参数初始化与数据加载，算力需求相对较低；收敛阶段是参数迭代更新的核心的阶段，计算与通信负载达到峰值，算力需求最高；微调阶段主要优化模型精度，计算与通信负载逐渐下降，算力需求趋于平稳，这种阶段性变化导致算力需求呈现动态波动[3]。（2）任务优先级变化：实际训练场景中，常存在多任务并行情况，任务优先级的动态调整会直接改变算力需求分配。例如，紧急任务插入时，需优先分配充足算力保障其快速推进，导致原有任务算力被调整；长尾任务处理过程中，随着任务推进，其算力需求会逐步变化，需动态适配，这些变化均构成了算力需求动态性的重要来源。

3 面向大模型训练的算力资源调度策略

3.1 整体架构设计

（1）分层调度模型：采用全局调度层与局部调度层协同工作的双层架构，明确各层职责、高效联动。全局调度层负责全局资源统筹分配，基于整个集群的资源状态（算力负载、网络带宽、内存占用）和所有训练任务的整体需求，制定资源分配全局策略，协调异构集群间的资源调度，避免全局资源冲突与浪费；局部调度层部署于各节点内部，负责接收全局调度指令，细化本地资源分配，实时响应节点内任务的动态变化，优化本地计算与通信的协同效率，弥补全局调度对局部节点细节适配不足的问题。（2）监控与反馈机制：以资源状态实时采集为核心，构建全流程、高精度的监控与动态反馈闭环。通过部署轻量化监控代理，实时采集集群内各设备（GPU/TPU/CPU）的算力利用率、内存带宽、网络延迟、能耗等核心指标，以及训练任务的计算负载、通信开销、参数更新进度等任务状态数据，采集频率适配训练动态特性，确保数据实时性；将采集到的数据汇总至调度中心，经分析处理后形成反馈信息，为分层调度决策、优先级调整、资源分配优化提供数据支撑，实现“监控-分析-反馈-调整”的动态闭环[4]。

3.2 动态优先级分配算法

（1）基于任务紧急程度、资源需求、历史表现的优先级模型：构建多维度优先级评价体系，综合量化任务优先级。以任务紧急程度（如项目截止时间、紧急迭代需求）作为核心权重，优先保障紧急任务的资源供给；结合任务资源需求（计算密集型/通信密集型、GPU核心数需求、内存带宽阈值），对资源需求匹配集群空闲资源的任务适当提升优先级，提升资源利用率；引入任务历史表现（如过往训练稳定性、资源利用效率、延迟率），对历史表现优良的任务给予优先级倾斜，对资源浪费严重、稳定性差的任务适当降低优先级，优化整体调度效益。（2）优先级动态调整策略：采用指数加权移动平均（EWMA）算法，实现优先级的动态自适应调整，避免优先级固化。基于实时监控的任务状态数据和资源状态变化，对任务优先级进行动态更新，赋予近期数据更高权重，快速响应任务紧急程度、资源需求的突发变化；当任务进入不同训练阶段（预热、收敛、微调）或出现资源需求波动时，自动调整优先级权重，同时设置优先级调整阈值，避免频繁调整导致的调度震荡，兼顾动态性与稳定性。

3.3 资源分配优化策略

（1）异构资源适配：针对GPU/TPU/CPU混合异构集群的特点，打破单一资源分配模式，实现多维度资源联合分配。结合训练任务的类型的（计算密集型优先分配GPU/TPU，轻量计算任务分配CPU），精准匹配GPU核心数，避免核心资源闲置；根据任务参数规模、数据量大小，分配适配的内存容量，保障参数存储与数据读写效率；同步匹配网络带宽资源，对通信密集型任务优先分配高带宽节点，实现GPU核心数、内存容量、网络带宽的协同适配，最大化异构集群的整体算力利用率。（2）通信-计算重叠优化：针对分布式训练中通信与计算相互制约的问题，采用通信-计算重叠策略，隐藏通信延迟，提升整体训练吞吐量。通过任务拆分与调度优化，将梯度同步等通信任务与前向传播、反向传播等计算任务并行执行，例如在部分节点进行前向传播计算时，同步开展其他节点的梯度聚合与参数同步操作；优化通信任务的执行时机，利用计算任务的间隙完成数据传输与参数同步，减少“计算等待通信”的时间损耗，提升通信与计算的协同效率[5]。

3.4 能耗感知调度

（1）动态电压频率调整（DVFS）与任务迁移结合：采用DVFS技术，根据训练任务的算力需求动态调整设备的电压与频率，避免高能耗闲置。在任务负载较低（如预热、微调阶段）时，降低设备电压与频率，减少能耗

损耗；在任务负载峰值（如收敛阶段）时，提升设备运行参数，保障计算性能；同时结合任务迁移策略，将低负载任务迁移至低能耗节点，将高负载任务集中分配至高性能节点，避免单一节点长期高负载运行导致的能耗浪费，实现能耗与性能的动态平衡。（2）碳感知调度：响应绿色计算需求，引入碳感知机制，结合可再生能源（太阳能、风能）的实时利用率调整调度策略。实时采集集群供电系统中可再生能源的占比与供电稳定性，当可再生能源利用率较高时，优先调度高能耗训练任务，充分利用清洁电力；当可再生能源供应不足时，调整任务执行顺序，优先执行低能耗任务，降低化石能源消耗，在保障训练进度的同时，减少碳排放，实现绿色高效调度。

4 讨论与未来工作

4.1 策略局限性分析

（1）对超大规模集群（万卡级）的扩展性不足：当前调度策略的分层架构的与资源分配逻辑，主要针对千卡级集群设计。当集群规模扩大至万卡级，全局调度层的资源状态采集、全局决策的计算开销会显著增加，易出现调度延迟；局部调度层与全局调度层的联动效率下降，可能导致资源分配冲突，难以实现万卡级集群的高效协同调度，扩展性有待提升。（2）动态优先级模型的参数敏感性较高：优先级模型的评价效果依赖于紧急程度、资源需求、历史表现等维度的权重参数设置。当前参数设置多基于经验校准，缺乏自适应调整能力，当训练任务类型发生较大变化（如密集型任务与轻量任务占比突变）时，参数敏感性会导致优先级评价偏差，进而影响调度公平性与效率，模型鲁棒性有待优化。

4.2 未来研究方向

（1）结合强化学习的自适应调度：引入强化学习算法，构建自适应调度模型，让调度系统通过与训练环境的持续交互，自主学习任务特征与资源状态的关联规律，自动优化优先级权重与资源分配策略，无需人工干预即可适配不同训练场景与任务类型，提升调度的智能化水平。（2）跨数据中心算力调度优化：针对多数据中心分布式训练场景，研究跨数据中心的算力调度策略，突破单数据中心的资源限制。优化跨数据中心的网络通信延迟与数据传输效率，实现多数据中心资源的全局统筹分配，提升算力利用率，保障跨域训练任务的稳定性。（3）量子计算与经典计算混合调度：随着量子计算技术的发展，探索量子计算与经典计算的混合调度模式。针对大模型训练中的特定计算任务（如复杂矩阵运算），适配量子计算设备的算力优势，合理划分量子计算与经典计算的任务边界，优化混合集群的资源分配，进一步提升训练效率、降低能耗。

结束语

本研究围绕大模型训练算力资源调度展开，提出分层调度架构、动态优先级分配、资源分配优化及能耗感知调度等策略，有效提升了算力利用率、训练效率并平衡了能耗。然而，面对超大规模集群和复杂任务场景，策略仍有优化空间。未来，将结合强化学习、探索跨数据中心及量子 - 经典混合调度，持续推动大模型训练算力调度技术的创新发展。

参考文献

[1]段晓东,李婕妤,程伟强,等.面向智算中心的新型以太网需求与关键技术[J].电信科学,2024,40(6):146-159.

[2]郭亮,王少鹏,权伟,等.面向大模型的智算网络发展研究[J].电信科学,2024,40(8):137-145.

[3]王学聪,冀思伟,李聪.面向大模型预训练的智算网络技术研究[J].电信科学,2024,40(12):160-172.

[4]梁小鸥.云网融合虚拟仿真实训室的建设研究[J].工程技术研究,2023,8(21):153-154.

[5]赵正波,王洁.面向算力的云网融合业务承载方式探讨[J].现代信息科技,2023,7(16):10-14.