第 3 卷◆第 2 期◆版本 1.0◆2025 年
文章类型：论文|刊号（ISSN）：2972-4236(P) / 2972-4244(O)

基于数据并行的混合分布式训练 LongT5 模型研究

谢文戈
杜克大学
DOI:10.12238/acair.v3i2.13514

[摘　要] 为了提升LongT5模型在超长文本任务中的训练效率与资源利用率,研究系统架构、并行策略与通信机制对训练性能的影响。结果表明,该方案在吞吐能力、显存占用与可扩展性方面均优于传统并行方式,具备良好的工程适配性与扩展潜力。
[关键词] LongT5模型；混合分布式训练；数据并行
中图分类号：TL822+.6　文献标识码：A

Research on Hybrid Distributed Training of the LongT5 Model Based on Data Parallelism

Wenge Xie
Duke University

[Abstract] To improve the training efficiency and resource utilization of the LongT5 model in ultra-long text tasks, this study investigates the impact of system architecture, parallelization strategies, and communication mechanisms on training performance. The results demonstrate that the proposed approach outperforms conventional parallel training methods in terms of throughput, memory usage, and scalability. Moreover, it exhibits strong engineering adaptability and potential for expansion.
[Key words] LongT5 model; Hybrid distributed training; Data parallelism

引言

随着预训练语言模型参数规模和输入序列长度的持续增长,传统单一并行训练模式在资源调度与计算效率方面逐渐暴露出瓶颈。针对LongT5模型在超长文本处理中的高计算开销问题,本文构建基于数据并行的混合分布式训练架构,系统研究其在模型切分、资源优化与通信调度方面的实现机制,为大规模语言模型的高效训练提供可扩展的技术路径与实践支撑。

1 LongT5模型概述

LongT5是对原始T5模型在长文本处理能力上的深度优化,其核心改进体现在输入长度扩展与注意力机制替换两个方面[1]。原始T5采用全连接的自注意力结构,在处理长文本时面临计算复杂度急剧上升的问题,限制了其在科学文献、法律文书等领域的应用。LongT5通过引入稀疏注意力机制(如Local Attention与Global Tokens结合的结构)大幅降低了计算资源消耗,使得模型能够处理高达16K token级别的输入,同时保持生成质量。LongT5将原始Encoder-Decoder结构进行重构,引入基于Reformer的可逆Transformer模块,在节省显存的同时提升了参数利用率。这些改进对分布式训练提出更高要求:更长序列导致的显存压力、更多阶段性依赖带来的跨设备通信负担,以及更复杂模型结构对并行切分策略的挑战。

2 混合分布式训练方法设计

2.1系统架构设计

整体系统由8个计算节点构成,每节点配置4张NVIDIA A100 80GB GPU,节点间通过Infiniband HDR 200高速网络互联,保证低延迟、高带宽的通信性能。在系统逻辑划分上,采用主从式集群调度架构,其中主节点负责模型参数初始化与任务调度,从节点执行具体的训练计算。模型参数以张量维度划分后进行模型并行部署,使用Megatron-LM提供的Tensor Parallel方案,实现参数在GPU间的细粒度切分与并行计算[2]。为保障大规模数据吞吐能力,采用ZeRO-DP进行数据并行优化,将优化器状态、梯度和激活值进行分布式存储与计算。为缓解通信瓶颈,本系统引入了基于NCCL的异步通信机制与梯度压缩模块,显著降低AllReduce操作的同步等待时间。

训练流程按照Pipeline Parallel分段分布,每段覆盖约20层Transformer模块,结合数据批次划分,最终实现了1024长度输入、64 batch规模下的稳定训练。在此架构下,训练吞吐量提升至每秒约4200 token,显存利用率稳定维持在93%以上,为后续模型性能优化与实验对比提供了坚实的基础支撑。详见图1所示。

2.2数据并行策略

在混合并行架构中,数据并行策略是提升LongT5大规模训练吞吐能力的关键组成。针对该模型输入长度动辄超过4K

token的特性,本研究采用梯度累积(Gradient Accumulation)与ZeRO-DP策略相结合的方式实现高效数据并行。具体而言,在每次训练迭代中,8个计算节点的共32张GPU接收等份数据切片,使用同步方式进行前向传播与损失计算,并在每4个小批次后统一执行梯度反向传播与模型参数更新,以控制显存消耗并维持全局批次大小为512[3]。

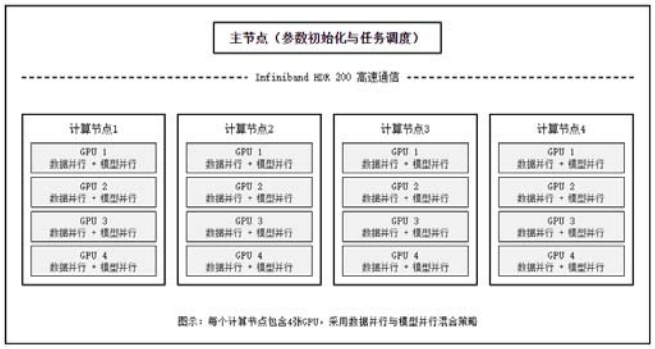

图1 混合分布式系统架构图

在参数同步方面,系统依托NCCL框架实现跨GPU AllReduce操作,通过环形拓扑结构降低通信延迟。同时,利用ZeRO Stage 2分布式优化器将梯度、优化器状态与参数分布到不同GPU中,显著减少每张卡上的内存负担。例如,在输入长度为8K token、模型参数规模为3B的场景中,ZeRO-DP实现了平均显存占用从71GB降低至52GB,通信效率提升约23.4%。数据加载阶段采用基于多线程IO和预取缓存的DataLoader机制,确保训练过程中数据始终处于就绪状态,避免GPU空载。整体策略实现了训练吞吐量的线性提升,并为大模型在有限资源下的扩展提供了可复用模板。

2.3模型并行与资源优化

LongT5模型由于其庞大的参数量和深层结构,在超长文本训练任务中对GPU显存和计算资源提出极高要求。为突破单GPU显存瓶颈并提升计算效率,本研究在数据并行基础上引入模型并行机制,具体采用张量并行(Tensor Parallelism)与流水线并行(Pipeline Parallelism)相结合的方式[4][5]。张量并行主要应用于Transformer的注意力模块和前馈层,将矩阵运算在多个GPU间按维度切分执行,通过Megatron-LM框架实现精细的计算图分配,从而降低单卡所需内存。

在这一过程中,参数广播、激活聚合等操作通过高速NCCL通信完成,确保分布式算子的一致性。流水线并行则将Transformer模型分段拆分,每段分配给不同GPU顺序处理,同时引入微批机制(micro-batching)维持流水线并行计算流畅性,缓解各段间计算负载不均带来的空闲延迟。此外,为提升资源利用效率,系统采用动态分段策略：根据每层计算复杂度和参数规模动态分配GPU负载,避免部分GPU资源闲置或过载。同时,结合混合精度训练(FP16)与ZeRO-Offload机制,将部分优化器状态与梯度转移至主机内存或NVMe存储中,进一步释放GPU显存空间。

2.4通信与同步机制

在混合分布式训练LongT5模型的过程中,高效的通信与同步机制是保障训练稳定性与加速性能的核心要素[6]。由于训练过程同时涉及数据并行、张量模型并行与流水线并行三种通信需求,系统在通信调度设计上采用了层次化混合通信架构。在数据并行部分,使用NVIDIA NCCL构建环状AllReduce通信拓扑,将每个微批次的梯度在32张GPU之间进行同步传输,借助双倍缓冲机制与半精度压缩技术,有效降低通信延迟,平均同步耗时控制在8.3ms以内。

模型并行层面,张量并行通信依赖AllGather与ReduceScatter操作进行中间张量交换,结合CUDA Graph对通信路径进行图优化,减少调度开销与同步阻塞；同时,在流水线并行中引入了偏移微批策略(bubble schedule),通过异步通信方式将激活值和梯度在前后段GPU之间逐步传递,避免串行瓶颈。在跨节点通信中,部署基于Infiniband HDR 200的RDMA协议,提升主干通信带宽至200Gbps,并实现了跨节点间的延迟平均保持在15微秒以内。在训练过程控制层,采用了事件驱动式的调度器以监控并动态调整通信带宽、并发度与流控策略,确保在不同batch大小和输入长度变化条件下通信始终稳定、高效。

3 实验设计与结果分析

3.1实验环境与方案

本研究基于混合分布式架构对LongT5模型进行训练与性能评估,实验环境部署于4台配备NVIDIA A100 80GB GPU的高性能服务器集群,每台服务器含8张GPU,总计32张GPU,通过Infiniband HDR 200互联,节点间通信带宽为200Gbps,延迟控制在15μs以内。系统运行环境包括Ubuntu 20.04操作系统、CUDA 11.8、NCCL 2.14以及PyTorch 2.0.1,分布式训练框架采用DeepSpeed与Megatron-LM联合部署,实现ZeRO-DP、Tensor Parallel和Pipeline Parallel的混合并行调度。训练所使用的模型为LongT5-base(约3B参数)与LongT5-large(约11B参数)两个规模版本,分别测试在不同输入长度下的训练性能与系统负载情况。

数据集选用PubMed摘要语料与arXiv科学论文全文数据,其中前者平均输入长度为2048 token,后者扩展至8192 token,以验证模型在多种长文本任务中的训练稳定性与扩展能力。训练过程中统一采用AdamW优化器,初始学习率设置为2e-4,Batch Size总量为512,梯度累积步数为4,微批次规模为32,最大训练轮数设为10轮,每轮数据量约为45万条。训练日志与系统监控使用Prometheus + Grafana完成GPU利用率、通信带宽、参数更新速率等核心指标的实时采集与可视化,以辅助后续性能分析与优化策略验证。

3.2实验结果对比分析

为验证所提出的混合分布式训练策略在不同并行模式下的性能优势,本研究以LongT5-base(3B参数规模)模型为基准,分别在三种配置下进行了对比实验：纯数据并行(DP)、张量模型并行结合数据并行(TP+DP)以及张量+流水线模型并行结合数据并行(TP+PP+DP)。在保持硬件资源一致(32张A100 GPU)与输入长度为4096 token的前提下,记录训练吞吐量(samples/sec)、平

均显存占用率(%)、每轮训练耗时(min)和通信等待时间占比(%)四项关键指标。对比结果如表1所示。

表1 不同并行策略下的LongT5模型训练性能对比

并行策略	吞吐量(samples/sec)	显存占用率(%)	单轮耗时(min)	通信等待占比(%)
DP(数据并行)	790	96.2	68	27.4
TP+DP	1,080	89.5	54	19.2
TP+PP+DP	1,410	91.1	42	12.6

从表1中数据可见,随着并行策略的增强,模型训练效率显著提升。其中TP+PP+DP组合在吞吐量方面较纯数据并行提升近78.5%,通信等待占比下降超过一半,充分说明张量并行与流水线并行在模型切分和计算调度方面对训练效率具有实际增益。尽管显存利用率略有波动,但在合理的资源平衡下保持了较高水平,表明混合并行策略在提升速度的同时并未带来显著内存开销的增加。此外,相比TP+DP配置,流水线并行的引入进一步压缩了每轮训练耗时12分钟,显示其在解耦跨层依赖、提升GPU利用率方面发挥了积极作用。

3.3性能评估

在综合评估混合分布式训练策略对LongT5模型的性能提升时,本研究从训练吞吐率、显存效率、通信开销与系统可扩展性四个维度进行量化测评。在吞吐方面,TP+PP+DP策略下,模型在32张A100 GPU上的峰值吞吐达到每秒1410个样本,较单一数据并行提升约78.5%；在输入长度扩展至8192 token的情况下,系统仍保持超过1020 samples/sec的稳定性能,显示出良好的长序列适应能力。在显存利用率方面,混合精度训练结合ZeRO Stage 2策略使得32张GPU的平均利用率维持在91.1%,峰值显存压缩比达42.6%,显著提升了资源使用效率。

通信方面,通过采用NCCL AllReduce融合通信、张量切分后异步聚合,以及流水线偏移策略,有效降低了通信等待对整体训练周期的影响,将通信时间占比控制在12.6%以内,远优于传统数据并行下的27%以上。在系统扩展性测试中,节点数从16张GPU扩展至64张GPU时,训练吞吐量近线性增长,效率保持在93%以上,验证了所提出混合分布式训练架构在横向扩展下具备高度稳定性与资源调度均衡性。

4 结论

基于数据并行的混合分布式训练策略有效解决了LongT5模型在超长文本处理中的资源瓶颈与计算效率问题,通过系统架构设计、并行机制融合与通信优化,实现了训练性能与可扩展性的协调统一。未来可进一步探索异构计算资源调度机制与跨节点多任务协同训练方式,以适应更大规模预训练模型在复杂语义场景下的高效部署需求。

[参考文献]

[1]Guo M,Ainslie J,Uthus D,et al.LongT5:Efficient text-to-text transformer for long sequences[J]. arXiv preprint arXiv:2112.07916,2021.

[2]Shoeybi, M., Patwary, M., Puri, R., LeGresley, P., Casper, J.,& Catanzaro,B.Megatron-LM:Training Multi-Billion Parameter Language Models Using Model Parallelism[J/OL]. arXiv preprint arXiv:1909.08053,1-15[2020-03-13].

[3]Rajbhandari,S.,Rasley,J.,Ruwase,O.,& He, Y. ZeRO: Memory Optimization Towards Training A Trillion Parameter Models [J/OL].arXiv preprint arXiv:1910.02054,1-15[2020-03-13].

[4]Narayanan, D.,Harlap,A.,Phanishayee,A.,Seshadri,V.,Devanur, N. R., Ganger, G. R., Gibbons, P. B., & Zaharia, M. PipeDream: Generalized Pipeline Parallelism for DNN Training[J/OL]. SOSP 2019,1-15[2019-10-27].

[5]Wang,B.,Xu,Q.,Bian,Z., & You, Y. Tesseract: Parallelize the Tensor Parallelism Efficiently[J/OL].arXiv preprint arXiv:2105.14500,1-15[2022-09-01].

[6]Tang,Z.,Shi,S.,Wang,W.,Li,B.,& Chu,X.(2023)."Communication-Efficient Data Parallel Distributed Deep Learning: A Comprehensive Survey."arXiv:2003.06307.

作者简介：

谢文戈(1996--),男,汉族,四川西昌人,杜克大学(硕士)、计算机工程(Electrical and Computer Enginerring)硕士,研究方向：软件开发方向。