

支持 Colossal AI 的训练在线监控工具创新赛 ——解决方案

团队名称：今天你科研了吗

一、团队介绍

本团队成员来自浙江大学感知技术与智能系统研究中心，由三名博士研究生和两名硕士研究生组成。长期致力于机器学习、模型轻量化、边缘计算等领域的研究，多篇研究成果在领域知名期刊和会议上发表，对基于昇腾框架的模型部署以及在线训练监控方法有着浓厚的兴趣以及扎实的技术储备。

二、对 Colossal AI 所提供的并行策略特性的理解

Colossal AI 是一个专为构建大型人工智能模型而设计的开源框架，支持多种并行策略，包括数据并行、模型并行、流水并行以及张量并行。其中数据并行策略将数据分割到多个设备上，每个设备拥有完整的模型副本，独立计算梯度并同步更新。模型并行策略将模型分割到多个设备上，每个设备负责模型的一部分计算。而流水并行将模型按层分割到多个设备上，每个设备负责一部分层的计算，数据在不同设备间流水传递。张量并行则将模型的张量操作分割到多个设备上，每个设备负责一部分张量计算。这些并行策略的结合使用可以显著提高大型模型训练的效率和经济性。

三、梯度监控解决方法的总体设计

为了实现对 Colossal AI 训练过程中各 rank 上权重梯度的监控，我们提出以下设计方案，主要包括监控功能设计、用户接口设计以及实现技术三部分。

3.1 监控功能设计

使能方法：通过训练脚本使能工具，通过 json 文件控制工具行为，指定监控对象，输出格式和统计信息等。

梯度监控：每个权重按名字独立统计，支持分别采集聚合前和聚合后的权重梯度，支持统计 norm、max、min、mean 信息以及梯度累积后的权重梯度采集。

并行策略支持：支持在开启并行策略时的梯度监控。确保监控工具在不同并行策略下的兼容性和稳定性。

轻量化：尽可能减少代码量，优化代码逻辑，采用高效的统计信息采集和存储机制，减少对训练过程的干扰。

3.2 用户接口设计

仿照 msprobe.monitor 在线监控的用户接口、配置和输出格式，方便用户集成和使用。支持通过配置文件或命令行参数进行监控配置。提供灵活的配置选项，满足不同用户的需求。

3.3 实现技术

使用业界主流语言 Python 实现监控工具，支持多模态理解和生成模型的梯度监控。支持重计算、分布式优化器等使用场景。确保监控工具在复杂训练场景下的稳定性和可靠性。

四、预期产出内容

预期产出包括：

- (1) 实现一个功能完善的梯度监控工具，支持 Colossal AI 训练过程中的权重梯度统计信息采集。
- (2) 提供详细的用户文档和使用示例。
- (3) 提供监控工具在并行策略下的性能测试报告，分析监控过程对训练性能的影响。
- (4) 提供持续的技术支持和维护。

五、时间规划

第一阶段（2024.12.08-2024.12.31）：

熟悉 Colossal AI 并行策略和具体使用，设计监控工具的总体架构和接口。

第二阶段（2025.01.01-2025.01.31）：

实现监控工具的核心功能，包括权重梯度统计信息采集和并行策略支持，调试用户接口，进行初步的性能测试和优化。

第三阶段（2025.02.01-2025.03.04）：

编写软件设计文档，对解决方法进行详细的阐述。完成指导程序功能使用的完整用户手册。在 llama2-7B 和 StableDiffusion 上测试并给出作品测试报告

第四阶段（2025.03.05-2025.03.12）：

进行全面的性能测试和优化，改进监控工具的用户接口和配置管理，进一步完善软件设计文档、用户手册以及测试报告的相关内容。