

# AIPerf: 大规模人工智能 算力基准测试程序

## *AIPerf: large-scale AI system benchmark*



翟季冬 (1981- ), 男, 博士, 清华大学计算机科学与技术系副教授, 主要研究方向为并行计算、性能评测、编译优化。

中图分类号 :TP3

文献标识码 :A

doi: 10.11959/j.issn.2096-0271.20210032

基金项目: 国家自然科学基金资助项目 (No.U20A20226)

Foundation Item: The National Natural Science Foundation of China(No.U20A20226)

随着我国近几年人工智能新基建与大规模计算系统的逐步落地,为了使大规模人工智能计算系统健康持续发展,需要一个能够有效地评价计算系统的人工智能算力的评价指标,而现有的测试方法无法满足这一需求。

目前,已有一些大规模计算系统的评测程序,例如Linpack是一个目前被广泛使用的高性能计算机双精度浮点运算性能基准评测程序,国际超算Top500榜单依据Linpack值来进行排名。然而典型的人工智能应用并不需要双精度浮点数运算,大部分人工智能训练任务以单精度浮点数或16位浮点数为,推理以Int8为主。各大企业、高校和相关组织也相继开发了各类人工智能算力基准评测程序,如MLPerf、Mobile AI Bench、DeepBench、AIIA DNN Benchmark,以及在双精度的Linpack基础上改成混合精度的HPL-AI等基准评测程序,但是这些程序并不适用于大规模和可变规模人工智能计算系统的评测。

为了填补大规模计算系统人工智能算力评测这一领域的空白,清华大学与鹏城实验室联合研制并开发了人工智能算力基准测试程序——AIPerf Benchmark(以下简称AIPerf)。AIPerf基于微软NNI(neural network intelligence)开源框架实现,以自动化机器学习(AutoML)为负载,使用网络态射(network morphism)网络结构搜索和树状结构Parzen估计(tree-structured Parzen estimator, TPE)方法超参搜索来寻找精度更高的神经网络结构和(或)超参数。用户可以通过配置文件指定AutoML的相关参数,如训练使用的批大小(batch size)、最大epoch数、学习率、最大搜索模型总个数、最长搜索总时间、最大同时搜索模型个数(并发数)等多个参数。

AIPerf的设计达到了以下4个关键的设计目标。

#### (1) 一个统一的分数

AIPerf能够报告一个分数,该分数可以作为被评测计算集群系统的评价指标。AIPerf目前的评价指标是Tops,即平均每秒处理的混合精度AI浮点操作数。使用一个而不是多个分数能方便进行不同机器的横向比较,且方便公众宣传。

#### (2) 可变的问题规模

人工智能计算集群往往有不同的系统规模,差异性体现在节点数量、加速器数量、加速器类型、内存大小等指标上。因此,为了适应各种规模的高性能计算集群,AIPerf能够使用AutoML调整问题规模来适应集群规模的变化,从而充分利用人工智能计算集群的计算资源来体现其算力。

#### (3) 具有实际的人工智能意义

具有人工智能意义的计算(如神经网络运算)是人工智能基准测试程序相较于传统高性能计算机基准测试程序的重要区别,也是其能够检测集群人工智能算力的核心所在。目前,AIPerf通过在ImageNet数据集上训练神经网络来运行计算机视觉应用程序;将来,计划将自然语言处理等其他人工智能任务加入AIPerf的评测范围。

#### (4) 包含必要的多机通信

网络通信是人工智能计算集群设计的主要指标之一,也是其庞大计算能力的重要组成部分。作为面向高性能计算集群的人工智能基准测试程序,AIPerf包括必要的多机通信(如任务的分发、结果的收集与多机训练),从而将网络通信性能作为最终性能的影响因素之一。

在第二届中国超级算力大会(ChinaSC2020)上,基于AIPerf大规模人工智能算力基准评测程序的国际人工智能性能算力排行榜首次发布。鹏城实验室研

制的基于ARM架构和华为加速处理器的鹏城云脑二主机以194 527 Tops的AIPerf算力荣登榜首,其性能是排名第二的联泰集群NVIDIA系统性能的12倍以上。

AIPerf基准评测程序还处于积极的开发和完善阶段,目前AIPerf正在应用负载、硬件适应性以及国际推广等方面大力推进。AIPerf目前只支持面向计算机视觉的

人工智能应用程序。为了更好地评估大规模智能系统在各个典型应用领域的性能,研究团队计划之后支持更多种网络搜索与训练算法以及评测数据集,同时支持更多类型的人工智能计算集群系统;此外,将积极推动AIPerf在国际上的影响力,将其打造成有国际影响力和公信力的大规模人工智能系统基准评测程序。 □