
科研团队攻克深度学习优化难题

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/38865.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

科研团队攻克深度学习优化难题。近日，北京交通大学数学与统计学院教授周声龙、罗自炎团队在国际顶级人工智能期刊《自然·机器智能》成功发表创新成果，为深度学习优化领域的技术突破提供了全新解决方案。

当前，基础大模型的快速发展正在全球范围内引发范式转变，深刻重塑各行各业的发展模式。但长期以来，训练这些大模型所采用的主流优化器均基于随机梯度下降算法开发，存在收敛速度慢、对收敛条件假设苛刻等固有局限，特别是在分布式环境中出现的数据异质性问题，给算法的理论分析与数值性能带来了巨大挑战，已成为资源受限下制约大模型高效训练与广泛应用的关键瓶颈。

针对这些挑战，研究团队经过持续攻关，提出了一种全新优化算法——预条件非精确随机交替方向乘子法（PISA）。该算法突破了深度学习优化领域收敛效率低、强假设依赖、计算复杂度高、泛化性不足四大核心瓶颈，为大模型训练提供了更高效稳健的技术路径。

该算法的创新之处在于，通过预条件化框架整合二阶信息、动量与正交化等技术，实现了高效并行计算，可适配各类大规模深度学习场景。与传统算法相比，PISA 算法展现出显著优势。在理论层面，在有界区域上梯度Lipschitz连续性这一弱假设下，实现线性收敛。因无需数据独立同分布、梯度有界、方差有界等强约束，从而有效处理数据异质性难题；在实践层面，其衍生变体SISA（二阶矩预条件）和NSISA（牛顿-舒尔茨正交化动量预条件）计算高效，泛化性强，在视觉模型、大语言模型、强化学习、生成对抗网络等多种模型架构的训练与微调中，收敛速度、精度与稳定性超越主流优化器。

上述理论优势与实践性能，在实验中得到了充分验证。例如，在利用MNIST数据进行分类任务中，当每个训练节点仅包含1个标签数据时（此时数据分布处于完全偏斜场景），算法准确率从54.33%提升至94.97%，大幅改善了数据异质性带来的性能损耗。在训练大语言模型时，随参数量增加，算法优势愈发明显，全量微调GPT2-XL（15亿参数）的时间至少缩短一半。在生成对抗网络训练中，SISA的测试FID值（衡量生成图像与真实图像接近程度的核心指标）为85.07，较主流的Adam算法（95.06）降幅约10.5%，生成效果显著提升。

周声龙表示，团队将继续深化研究，推动算法成果的产业化转化，为我国人工智能产业高质量发展注入新动能，助力新质生产力发展进程中基础研究与应用研究的深度融合。（来源：中国科学报 王瑞霞 陈彬）

相关论文信息：<https://doi.org/10.1038/s42256-026-01182-3>

作者：周声龙等 来源：《自然—机器智能》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发