arXiv:2303.15053v1 [physics.data-an] 27 Mar 2023

使用高性能计算对基于人工智能的高能物理工作负载进行超参数优化、量子辅助模型性能预测和基准测试

Eric Wulff、MariaGirone、David Southwick、Juan Pablo Garc´ıa Amboage 和 Eduard Cuba
欧洲核子研究组织，粒子大道1号，日内瓦1211号23区，瑞士。电子邮箱：eric.wulff@cern.ch

摘要。 基于深度学习的人工智能模型的训练和超参数优化（HPO）通常需要大量计算资源，因此需要使用大规模分布式资源，以及可扩展且资源高效的超参数搜索算法。本研究探讨了利用模型性能预测辅助在高性能计算系统上开展的超参数优化流程的潜力。此外，我们使用量子退火器训练性能预测器，并提出一种方法，以克服当前量子系统局限性所带来的部分问题，同时提高解的稳定性。这使得我们能够在量子机器上取得与经典机器相当的结果，并展示了量子计算机如何集成到经典机器学习调优流程中。

此外，文中还介绍了一项容器化基准测试的开发成果。该基准测试基于用于碰撞事件重建的 AI 模型，可用于比较和评估不同硬件加速器在训练深度神经网络方面的适用性。

1. 引言

当前最先进的超参数（HP）搜索算法，如 Hyperband [1], 、异步逐次减半算法（ASHA）[2] 和贝叶斯优化Hyperband（BOHB） [3] ，都依赖提前终止机制：自动终止表现不佳的试验，以释放计算资源供新试验使用。

根据所选指标（通常是准确率或验证损失）对试验进行相对排名，并以此为依据设定停止准则，可能会带来问题。由于训练过程是非线性的，决策时点的试验排名未必能保持到目标时点，参见图1。解决这一问题的一种可能方法是采用非线性停止准则，例如使用支持向量回归（SVR），根据部分训练的模型预测最终模型性能 [4]。

大型高性能计算（HPC）系统尤其适合运行 ASHA 等超参数优化（HPO）算法，因为它们能够实现超线性扩展；这一点已在于利希超级计算中心（JSC）的 JURECA-DC [5]系统上进行的测试中得到验证，如图2所示。

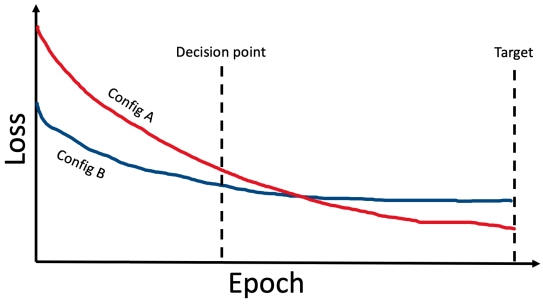


图1：不同超参数配置产生的两条学习曲线示意图。

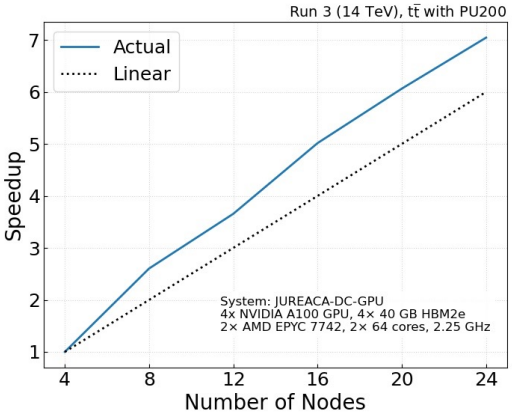


图2：HPO 工作流在 JURECA-DC-GPU 系统上的扩展情况。

2. 用于预测模型性能的量子支持向量回归 $\mathbf{n}$

本文研究了通过性能预测加速 HPO 流程的潜力，以及使用量子退火机（QA）训练性能预测器的可行性。量子支持向量回归（QSVR）的性能与经典支持向量回归相当，这表明量子计算机有望在未来集成到经典机器学习调参工作流中。

一种基于图神经网络（GNN）的算法，专为高能物理（HEP）中的机器学习粒子流（MLPF）重建任务开发 [6, 7]，是本研究的基础模型。通过在公开可用的 Delphes 数据集 [8]上训练 296 种不同的 MLPF 配置，生成了由学习曲线和超参数配置组成的数据集，见图3。这些配置从表1所定义的超参数空间中随机抽取。训练采用分布式数据并行模式，在每个配备四块 NVIDIA A100 GPU 的计算节点上运行。在这样的节点上训练一个模型 100 轮，大约需要 16 小时的实际运行时间。为了加快学习曲线的生成速度，训练在于利希超级计算中心（JSC）的 JURECA-DC-GPU 系统上并行使用了 24 个计算节点。

超参数	样本分布
学习率	对数 $lr \sim \mathrm{U}[10^{-6}, 3 \cdot 10^{-2}]$
分类图层	$\{0, 1, 2, 3, 4\}$
回归图层	$\{0, 1, 2, 3, 4\}$
丢弃率	$U[0.0, 0.5]$
分箱大小	$\{8, 16, 32, 64, 128\}$
输出维度	$\{8, 16, 32, 64, 128, 256\}$
权重衰减	对数 $wd \sim \mathrm{U}[10^{-6}, 10^{-1}]$

表 1：用于创建学习曲线数据集的超参数空间。

为建立基线，我们通过在不同的训练/测试划分上拟合 1000 个模型来研究经典支持向量回归，并对数据集大小施加与量子支持向量回归相同的约束。由于量子退火机上的量子比特数量有限，拟合时只能使用 20 个训练样本。结果因训练划分而异，统计量见图 4 和表 2。

通过欧洲百亿亿次计算卓越中心“面向百亿亿次计算的人工智能与仿真工程研究”（CoE RAISE），获得了使用于利希超级计算中心的量子退火机 JUPSI 的权限，

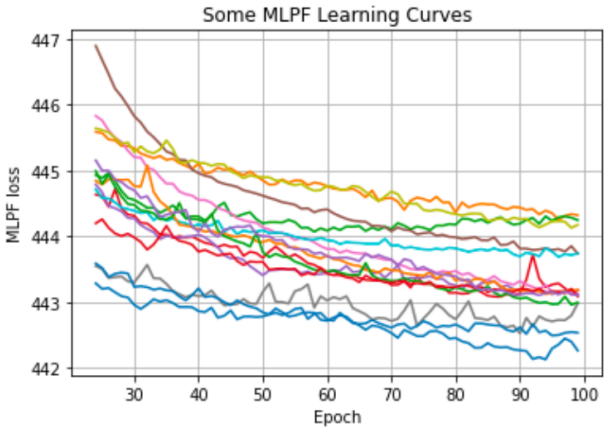


图 3：学习曲线数据集中表现最佳的学习曲线，放大显示第 25 至第 100 个训练轮次。

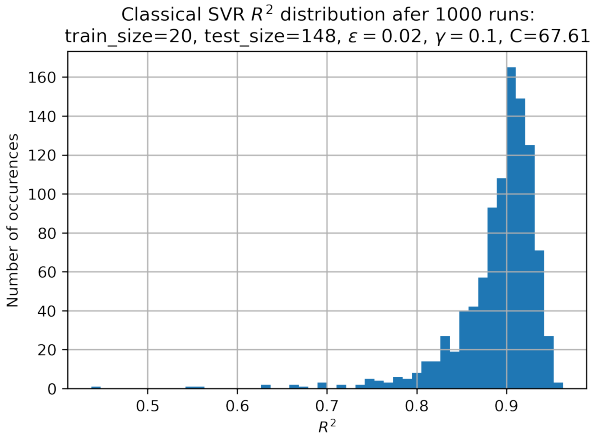


图 4：在经典计算机上对 1000 个不同划分进行支持向量回归训练和测试后得到的 R^2 分布。

并利用该设备训练量子支持向量回归模型，以进行模型性能预测。由于量子过程具有概率性，退火会运行多次，并返回多个解，之后可通过不同方式将这些解组合起来，生成最终的量子支持向量回归模型。 [9]介绍了组合量子支持向量回归解的不同方法。图 5 将性能最佳的量子支持向量回归模型的预测值与真实值进行了对比。

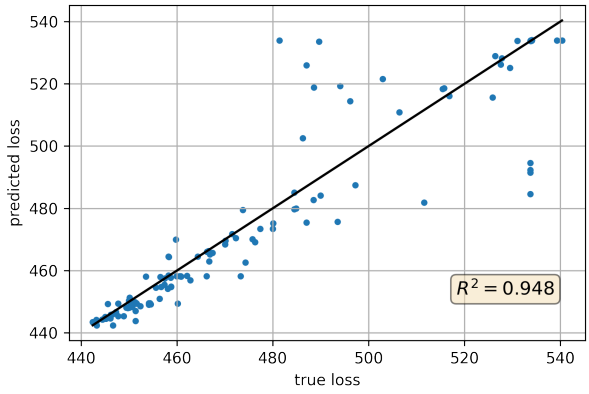


图 5：单个表现最佳的量子支持向量回归模型预测的验证损失值与真实验证损失值的对比。

为稳定量子支持向量回归性能，我们测试了不同方法，将多个分别使用 20 个样本训练的量子支持向量回归模型组合起来。我们发现，最成功的方法是将训练集划分为互不相交的、各含 20 个点的子集，并针对每个子集训练一个量子支持向量回归模型。最终的集成预测通过对所有单独预测取平均值计算得出。

为评估 QSVRcombination 技术，我们使用了 80 个训练点和 150 个测试点。该实验针对 10 种随机训练/测试划分重复进行，结果如图 6 所示。请注意，每次预测均由 4 个各自在 20 个点上训练的量子支持向量回归组合而成。尽管这种方法未能提高最高的 R^2 分数，但它显著改善了表现最差的划分结果，并降低了不同划分之间 R^2 分数的标准差，因此结果更加稳定，如表 2 所示。这一点值得关注，因为使用 Q-SVR 时面临的问题之一就是结果不稳定。还可以通过调整加权平均中的权重来进一步改进这些量子支持向量回归的组合方式；目前使用的是简单平均。表 2 展示了本文所述不同模型的 R^2分数统计量对比。

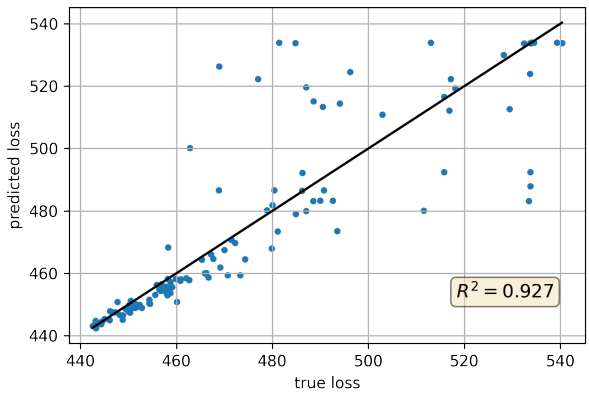


图 6：使用 80 个训练样本，针对 10 种不同的训练/测试划分运行 QSVRcombination 技术 10 次后，表现最佳的量子支持向量回归集成模型的预测值与真实值对比。

	Best	最差值	Mean	Std	数量 训练次数
SVR	0.959	0.318	0.889	0.050	1000
模拟量子支持向量回归	0.949	0.383	0.901	0.045	100
QSVR	0.948	0.742	0.880	0.056	10
量子支持向量回归集成	0.927	0.857	0.899	0.019	10

表 2： R^2 不同回归模型的数值和统计量。SVR 模型是在经典计算机上训练的经典支持向量回归模型。Sim-QSVR 模型是在经典计算机上使用模拟量子退火训练的量子支持向量回归模型。QSVR 模型是在量子退火机上训练的量子支持向量回归模型；量子支持向量回归集成则是由 4 个量子支持向量回归模型给出的预测结果组合而成，这些模型均在量子退火机上使用各 20 个训练样本进行训练。

3. 容器化人工智能基准测试

基于人工智能的算法有望在推理时提高准确率和重建可扩展性，但这需要先进行资源密集型训练。本文提出了一种自包含、可复现的容器化基准测试，该测试基于深度学习模型的训练，旨在探索将人工智能驱动的高能物理应用部署到不同高性能计算环境中的可行性。图 7 所示的指标是训练吞吐量，即每秒处理的训练样本数；测试使用的是公开可用的 Delphes 数据集子集 [8]，格式为 HEPscore [10]。

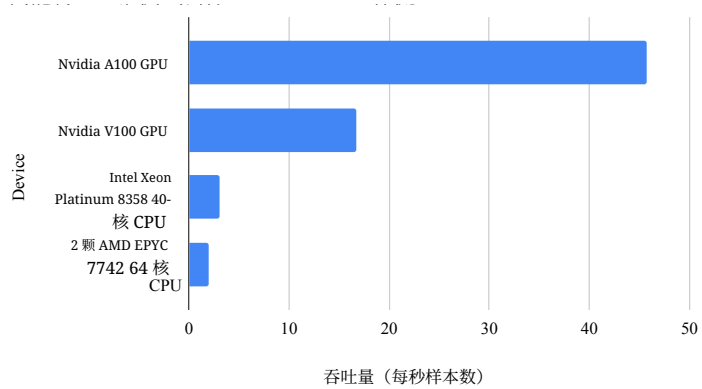


图 7：基准测试结果。

4. 结论

高性能计算系统对于运行大规模超参数优化和分布式训练至关重要，能够显著提升模型性能，并加快模型开发和训练的迭代速度。研究表明，超参数优化工作流可以实现超线性加速，体现了高性能计算在此类应用中的优势。

结果表明，利用性能预测进行超参数优化具有巨大潜力，这鼓励了未来的超参数优化研究采用这一技术。研究还表明，尽管量子计算机目前存在局限性，仍然可以在量子退火机上训练量子支持向量回归模型，并达到与经典支持向量回归相当的预测性能。这为进一步研究在超参数优化及其他应用场景中使用混合量子／高性能计算工作流提供了动力。

最后，开发一个容器化的基准测试应用，并采用一个来自高能物理的人工智能用例，可按照 HEPScore 格式快速、轻松地对新型硬件加速器进行基准测试。

致谢

我们感谢 CoE RAISE 的同事，尤其是 Andreas Lintermann 和 Marcel Aach，在本研究过程中提供了有益的讨论和反馈。我们还感谢 CMS 合作组的同事，特别是 Joosep Pata、Javier Duarte 和 Farouk Mokhtar，感谢他们在机器学习粒子流研究中的合作。

Eric Wulff、David Southwick 和 Eduard Cuba 的工作得到了 CoE RAISE 的支持。CoE RAISE 项目获得了欧盟“地平线 2020——研究与创新框架计划”H2020-INFRAEDI-2019-1 的资助，资助协议编号为 951733。

作者衷心感谢通过 JARA 获得的 JURECA 超级计算机使用时长，该计算机位于于利希研究中心 [5] 。作者衷心感谢于利希超级计算中心通过在 D-Wave AdvantageTM 系统 JUPSI 上提供 J¨ulich UNified Infrastructure for Quantum computing (JUNIQ) 的计算时长，为本项目提供资助。

参考文献

[1] Li L、Jamieson K、DeSalvo G、Rostamizadeh A 和 Talwalkar A，2016，Hyperband：一种用于超参数优化的新型基于多臂老虎机的方法，网址 https://arxiv.org/abs/1603.06560[2] Li L、Jamieson K G、Rostamizadeh A、Gonina E、Hardt M、Recht B 和 Talwalkar A，2018，大规模并行超参数调优 *CoRR* **abs/1810.05934** (*Preprint* 1810.05934)[3] Falkner S、Klein A 和 Hutter F，2018，BOHB：大规模场景下稳健高效的超参数优化*CoRR* **abs/1807.01774(***Preprint* 1807.01774)[4] Baker B、Gupta O、Raskar R 和 Naik N，2017，利用性能预测加速神经架构搜索，网址 https://arxiv.org/abs/1705.10823[5] 于利希超级计算中心，2021，JURECA：于利希超级计算中心采用模块化超级计算架构实现的数据中心型和加速器模块 *Journal of large-scale research facilities* **7**网址 http://dx.doi.org/10.17815/jlsrf-7-182[6]Pata J、Duarte J、Vlimant J、Pierini M 和 Spiropulu M，2021，MLPF：利用图神经网络进行高效的机器学习粒子流重建 *The European Physical Journal C* **81** 网址 https://doi.org/10.1140%2Fepjc%2Fs10052-021-09158-w[7] Pata J、Duarte J、Mokhtar F、Wulff E、Yoo J、Vlimant J、Pierini M 和 Girone M，2022，CMS 中用于粒子流重建的机器学习，网址 https://arxiv.org/abs/2203.00330[8] 等，J P，2021，使用 PYTHIA8+DELPHES3 模拟的 $t\bar{t}$ 和 QCD 粒子级事件，PU200，用于机器学习粒子流（MLPF），网址 https://zenodo.org/record/4559324[9] Pasetto E、Riedel M、Melgani F、Michielsen K 和 Cavallaro G，2022，利用遥感进行水体叶绿素浓度估算的量子支持向量回归 *IEEE Geoscience and Remote Sensing Letters***19** 1–5[10] Giordano D *et al.* 2021 WLCG 计算资源的 HEPiX 基准测试解决方案 *Comput. Softw. Big*$^{Sci.}$ **5** 28