追逐低碳电力，实现实用且可持续的深度神经网络训练

Zhenning Yang、Luoxi Meng、Jae-Won Chung、Mosharaf Chowdhury密歇根大学{znyang,luoxim,jwnchung,mosharaf}@umich.edu

arXiv:2303.02508v2 [cs.LG] 2 Apr 2023

摘要

近年来，深度学习发展迅速，使用图形处理器训练深度神经网络（DNN）导致能源消耗和碳排放增加。为响应可持续发展的呼吁，传统解决方案尝试将训练任务转移到碳强度较低的位置或时间段。然而，由于数据集规模庞大或受到数据法规限制，将任务转移到其他位置未必可行。此外，推迟训练可能会对应用服务质量造成负面影响，因为支撑服务的深度神经网络无法及时更新。本文提出了一种实用方案，无需迁移或推迟训练作业即可减少深度神经网络训练的碳足迹。具体而言，该方案会监测训练期间实时变化的碳强度，并控制图形处理器的能源消耗，从而在保持训练性能的同时减少碳足迹。此外，为主动适应碳强度的变化，我们提出了一种轻量级机器学习算法，用于预测即将到来的时间段的碳强度。我们的方案 *Chase* 将 ImageNet 上 ResNet-50 训练的总碳足迹降低了 13.6%，而训练时间仅增加 2.5%。

1 引言

深度学习的发展导致用于训练深度神经网络的图形处理器能耗和碳排放显著增加（Anderson et al.,2022;Wuetal.,2021;Patterson et al., 2021），提高深度神经网络训练的碳效率已成为一项迫切而紧急的问题。具体而言，训练 GPT-3（Brown et al., 2020）等大型深度神经网络会产生 552 公吨 CO_2 排放（Patterson et al., 2021）。
然而，并非所有焦耳都生而平等；碳强度是衡量发电情况的指标，计算方法是统计每千瓦时（kWh）电力产生的二氧化碳排放克数（g $\cdot CO_2$/kWh）。碳强度会因时间和位置不同而显著变化。例如，严重依赖煤炭发电的地区，其碳强度会高于依赖核能、太阳能或风能等无碳能源的地区（Miller et al., 2022; Google, 2018）。此外，碳强度还会随一天中的时段或季节而变化，因为许多可再生能源依赖自然现象。

在这项工作中，我们证明，通过预测并利用实时碳强度的变化，可以提高深度神经网络训练的碳效率。也就是说，当碳强度上升时，我们会放慢训练速度，以减少电力消耗；另一方面，当碳强度下降时，我们会加快训练速度，以取得更多进展。*Chase*会自动做出这些决策，在略微延长训练时间的同时，大幅减少碳排放。Chase 将开源。

2 相关工作

碳感知作业调度利用碳强度随时间（Li 等，2016；Haghshenas 等，2022）和位置（Moghaddam，2014；Berl 等，2009）的变化，以减少

深度神经网络训练的碳排放。然而，由于大型数据集（Caesar 等人，2019；Dai 等人，2017；Deng 等人，2009）、数据法规（《通用数据保护条例》，2018年）以及资源可用性等各种限制，将作业迁移到更绿色的地理位置并不总是可行。此外，将训练作业推迟到更绿色的时间也未必可行，因为深度神经网络必须使用最新数据进行训练，并快速部署到生产环境，以实现最高的服务质量。相比之下，我们的解决方案不会迁移或推迟训练作业。相反，在训练任务按请求运行的同时，我们会透明地调整其速度和能源消耗，使其自动追逐低碳电力。

优化深度神经网络训练的能源消耗，自然能够降低碳排放，因为碳排放与能源消耗呈线性关系。图形处理器是训练深度神经网络的主要硬件，用户可以通过软件设置其功率上限（英伟达，2022年）。利用这一技术，Zeus（You 等人，2023）通过自动配置多个重新训练任务的功率上限和批量大小，联合优化达到目标验证准确率所消耗的能源和时间。然而，Zeus 关注的是训练作业的时间和能源消耗，并未考虑碳强度及其随时间变化的特性。

要主动应对碳强度的变化，就需要获得下一个时间窗口的碳强度预测。近期方法（Maji 等人，2022a;b）已取得较高的预测性能，但使用深度神经网络会消耗图形处理器资源，并可能抵消后续优化技术所带来的部分碳足迹减排。另一方面，也有商业服务（WattTime，2022年；ElectricityMaps，2022年）提供历史碳数据和预测。然而，其高级预测功能的费用可能并非所有人都负担得起。我们认为，需要一种轻量、低成本的短期碳强度预测方案，以推动碳感知深度神经网络训练的普及。

3 方法

在本研究中，我们提出了一种切实可行的方法，以减少深度神经网络训练的碳足迹。我们根据碳强度的变化调整图形处理器的功率上限，从而同时优化碳排放和训练性能，本质上是优先选择低碳能源。此外，为了以经济实惠的方式准确预测未来时间窗口的碳强度，我们使用训练任务开始时间之前的历史碳强度数据，并拟合一个轻量级回归模型。

3.1 碳强度预测

训练期间，我们旨在通过预测直到下一次调用时的碳强度，定期调整图形处理器的功率上限。为了构建短期碳强度预测模型，在提交 DNN 训练任务时，我们会获取开始时间前一天的历史碳强度数据（T 个时间步），并拟合以下回归模型

$$\texttt{CarbonIntensity}(t) = f(\texttt{sin_time}(t),\ \texttt{cos_time}(t),\ \texttt{CarbonIntensity}(t-1)) \quad (1)$$

其中

$$\texttt{sin_time}(t) = \sin\left(\frac{2\pi \cdot t}{T}\right), \quad \texttt{cos_time}(t) = \cos\left(\frac{2\pi \cdot t}{T}\right). \quad (2)$$

我们的回归模型体现了这样一种直觉：下一时间步的碳强度不仅取决于当前碳强度，也取决于当前时刻，因为当前时刻会影响能源结构。此外，将时间步 t 转换为 sin time(t) 和 cos time(t)，以捕捉昼夜碳强度趋势的周期性。

用户可以配置要收集的历史数据量，以及预测与功率限制调整之间的间隔。例如，较短的间隔可以实现更精细的功率上限调节，但也会更频繁地进行预测。

3.2 碳感知 DNN 训练

在本节中，我们将开发一种在线优化算法，以适应不断变化的碳强度，并调整 p 的功率上限

图形处理器

以适应不断变化的碳强度。

深度神经网络训练的性能通常以达到精度的时间（TTA）来衡量，即达到给定目标精度所耗费的时间（Coleman 等，2019 年）。我们将整个过程中的碳排放定义为

这一过程的碳排放-精度 （CTA）：

$$\texttt{CTA} = \texttt{TTA} \times \texttt{AvgPower} \times \texttt{AvgCarbonIntensity} \tag{3}$$

其中，AvgPower 和 AvgCarbonIntensity 分别表示训练期间的平均功率和平均碳强度。我们可以将问题表述为一个随时间变化的成本最小化问题，其中成本可定义为：

$$\eta \cdot \texttt{CTA} + (1-\eta) \cdot \texttt{MaxPower} \cdot \texttt{MaxCarbonIntensity} \cdot \texttt{TTA} \tag{4}$$

其中， $\eta \in [0,1]$ 是一个可配置参数，用于预先指定碳效率与训练性能的相对重要性。GPU 最大功率和最大碳强度是常数，用于统一成本指标（gCO_2）中的计量单位，并平衡这两项。

将公式 3 代入公式 4，可得以下成本表达式：

$$\text{TTA} \cdot (\eta \cdot \text{平均功率} \cdot \text{平均碳强度} + (1-\eta) \cdot \text{最大功率} \cdot \text{最大碳强度}) \tag{5}$$

由于难以准确刻画公式 5 中的两个项，直接求解完整的最小化问题较为困难：

> 1. 平均碳强度：虽然短时间内的碳强度可能可以预测，但要可靠地预测整个训练周期内的碳强度却很困难，而训练可能持续数天甚至数周。2. TTA：深度神经网络训练具有随机性，因此很难预测 TTA。

我们的见解是，碳强度在短时间内会相对稳定，因此可以按时段优化成本。为此，我们提出以在线方式迭代优化成本：从时间步$t \in [1,T]$开始，预测每个时段 的碳强度（§3.1），并在每个时段开始时确定最优 GPU 功率上限 p。因此，对于每个时段，我们求解以下优化问题：

$$\min_{p \in \mathcal{P}} \frac{\eta \cdot \texttt{AvgPower}(p) \cdot \texttt{CarbonIntensity}(t) + (1-\eta) \cdot \texttt{MaxPower} \cdot \texttt{MaxCarbonIntensity}}{\texttt{Throughput}(p)} \tag{6}$$

其中，AvgPower(p) 表示设定功率上限 p 时经分析得到的功耗，Throughput(p) 与 TTA 成反比，因为改变 GPU 的功率上限不会改变模型将训练的样本数量。我们的公式受到 Zeus（You 等人，2023）的启发，但不同之处在于，我们纳入了实时碳强度，并根据其变化进行调整。

总而言之，当训练任务到达时，我们的系统首先针对允许的 GPU 功率上限集合中的所有 p，对吞吐量(p)和平均功率(p)进行性能分析 $\mathcal{P}$。用户可以指定每个周期的时长，这决定了训练过程中优化成本的频率。每个周期开始时，我们会预测碳强度(t)，并通过求解公式 6 来确定该周期的最优功率上限。通过定期重新评估，我们能够优化整个过程的总体成本，并实现碳高效的深度神经网络训练。

4 结果

4.1 预测性能

在评估中，我们发现，当检索和预测的时间间隔短于 10 分钟时，碳强度的平均变化幅度不到 0.1%。因此，我们使用 WattTime API（WattTime，2022年）检索了美国中部地区从 2023-01-15 到 2023-01-27（格林威治标准时间）的历史碳强度轨迹，时间间隔为 30 分钟（共 552 个数据点）。

我们发现，DNN 训练任务开始前 24 小时的数据足以拟合回归模型。其余 504 个数据点（即 252 小时的数据）用于测试。我们测试了一系列

回归模型（表 1）。为了评估这些模型的性能，我们采用了时间序列预测中常用的平均绝对百分比误差（MAPE）指标。支持向量回归（SVR）（Basak 等人，2007）是用于碳感知 DNN 训练时表现最佳的模型。

表 1：碳强度预测模型性能对比。

模型	MAPE %
支持向量回归	**0.94**
线性回归	1.57
梯度提升	2.23
自适应提升	2.51
随机森林	1.76

4.2 深度神经网络训练

为评估我们方案的有效性，我们使用一块英伟达 A40 GPU，在 ImageNet 数据集（Deng 等人，2009）上训练 ResNet50（He 等人，2015）。MaxPower 设置为 300W，这是 A40 GPU 所支持的最高功率上限。MaxCarbonIntensity 设置为 750g $\cdot CO_2$/kWh，即训练任务开始时间之前 24 小时内观测到的最大碳强度。我们将该方法与常规训练进行比较；常规训练使用默认的 GPU 配置（即使用 MaxPower）运行相同任务。

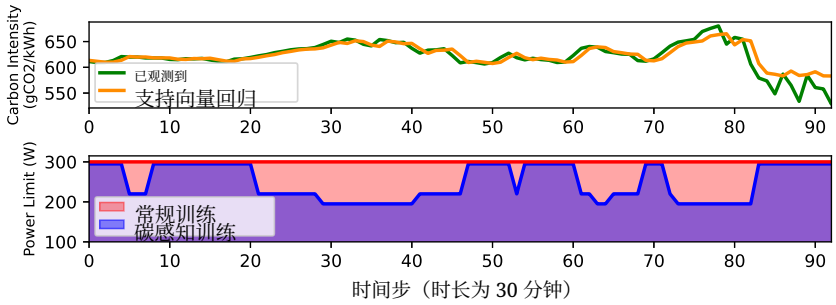

图 1：在训练过程中，功率上限会动态调整，以适应碳强度的波动。A40 GPU 的默认功率上限为 300W。使用默认 GPU 配置进行训练会导致更高的能源消耗，进而产生更多的碳排放。

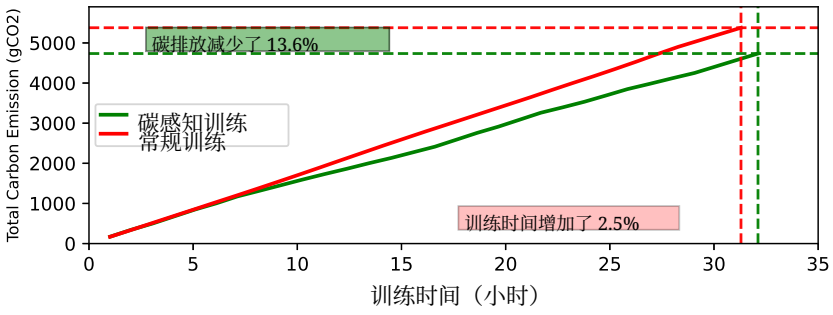

图 2：与常规训练相比，碳感知训练在整个训练过程中都能减少碳排放，并以略长的训练时间达到相同的准确率。

与常规深度神经网络训练方法相比，我们的解决方案有效地将总碳足迹减少了 13.6%（图 4.2）。这是通过减少电力使用并调整动态功率实现的

上限以优先使用更绿色的能源（图 4.2），训练时间仅增加 2.5%，因此即使是对时间要求严格的 DNN 训练任务，也能立即减少碳排放。

5 结论

总之，本研究致力于解决如何降低图形处理器上深度神经网络训练的能源消耗和碳排放问题。通过采用简单的回归模型和少量历史数据，我们展示了该方法能够实现较高的短期预测性能。利用这些信息，我们的解决方案能够实时动态、自动地调整图形处理器功率上限，在无需作业迁移或延期的情况下减少碳排放。作为未来工作，我们认为，扩展 Chase 以支持数据中心中的多个 DNN 训练任务，将有助于显著推动应对气候变化的工作。

6 致谢

我们谨向审稿人和 SymbioticLab 成员致谢，感谢他们提出的宝贵意见。本研究部分由美国国家科学基金会的 CNS-1909067 和 CNS-2104243 项目以及 VMware 的一项资助支持。Jae-WonChung 还获得了宽正教育基金会的资助。

参考文献

Thomas Anderson、Adam Belay、Mosharaf Chowdhury、Asaf Cidon 和 Irene Zhang。Treehouse：碳感知数据中心软件的案例，2022年。网址：https://arxiv.org/abs/2201. 02120。

DebasishBasak、SrimantaPal 和 Dipak Patranabis。支持向量回归。《神经信息处理：快报与评论》，11，11 2007年。

Andreas Berl、Erol Gelenbe、Marco Di Girolamo、Giovanni Giuliani、Hermann DeMeer、Minh Quan Dang 和 Kostas Pentikousis。高能效云计算。《计算机杂志》，53（7）：1045–1051，08 2009年。ISSN 0010-4620。doi：10.1093/comjnl/bxp080。网址：https://doi. org/10.1093/comjnl/bxp080。

Tom B. Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared Kaplan、Prafulla Dhari- wal、Arvind Neelakantan、Pranav Shyam、Girish Sastry、Amanda Askell、Sandhini Agarwal、Ariel Herbert-Voss、Gretchen Krueger、Tom Henighan、Rewon Child、Aditya Ramesh、Daniel M. Ziegler、Jeffrey Wu、Clemens Winter、Christopher Hesse、Mark Chen、Eric Sigler、Mateusz Litwin、Scott Gray、Benja min Chess、Jack Clark、Christopher Berner、Sam McCandlish、Alec Radford、Ilya Sutskever 和 Dario Amodei。语言模型是少样本学习者。《CoRR》，abs/2005.14165，2020年。网址：https://arxiv.org/abs/2005.14165。

Holger Caesar、Varun Bankiti、Alex H. Lang、Sourabh Vora、Venice Erin Liong、Qiang Xu、Anush Krishnan、Yu Pan、Giancarlo Baldan 和 Oscar Beijbom。nuScenes：面向自动驾驶的多模态数据集。*CoRR*，abs/1903.11027，2019 年。网址：http://arxiv.org/abs/1903. 11027。

Cody Coleman、Daniel Kang、Deepak Narayanan、Luigi Nardi、Tian Zhao、Jian Zhang、Peter Bailis、Kunle Olukotun、Chris R´e 和 Matei Zaharia。DawnBench 分析：一项以达到精度所需时间为指标的机器学习性能基准。*SIGOPS* 操作系统评论，53(1):14–25，2019 年 7 月。ISSN 0163-5980。doi：10.1145/3352020.3352024。网址：https://doi.org/10.1145/ 3352020.3352024。

Angela Dai、Angel X. Chang、Manolis Savva、Maciej Halber、Thomas Funkhouser 和 Matthias Nießner。《ScanNet：室内场景的丰富标注三维重建》。发表于 计算机视觉与模式识别会议（*CVPR*）论文集，*IEEE*，2017 年。

JiaDeng、Wei Dong、Richard Socher、Li-Jia Li、Kai Li 和 Li Fei-Fei。ImageNet：一个大规模分层图像数据库。载于*2009*年 *IEEE*计算机视觉与模式识别会议，第 248–255 页，2009 年。doi：10.1109/CVPR.2009.5206848。

ElectricityMaps。电力地图，2022年。网址：https://www.electricitymaps.com/。

《通用数据保护条例》。2018年《通用数据保护条例》第5条：个人数据处理原则。网址：https: // gdpr-info.eu/art-5-gdpr/。

谷歌。谷歌数据中心迈向全天候无碳能源：进展与见解，2018年。网址：https://www.gstatic.com/gumdrop/sustainability/ 24x7-carbon-free-energy-data-centers.pdf。

Kawsar Haghshenas、Brian Setz 和 Marco Aiello。面向深度神经网络训练工作负载的二氧化碳排放感知调度。载于*2022*年*IEEE*大数据国际会议（大数据），第1542–1549页，2022年。doi：10.1109/BigData55660.2022.10020544。

何恺明、张祥雨、任少卿和孙剑。用于图像识别的深度残差学习。*CoRR*，abs/1512.03385，2015年。网址：http://arxiv.org/abs/1512.03385。

李超、王睿、钱德沛和李涛。基于可再生能源组合管理服务器集群。*ACM*自主与自适应系统汇刊，11(1)，2016年2月。ISSN 1556-4665。doi：10.1145/2845085。URL https://doi.org/10.1145/2845085。

Diptyaroop Maji、Prashant Shenoy 和 Ramesh K. Sitaraman。Carboncast：电网碳强度的多日预测。载于第九届*ACM*能源高效建筑、城市与交通系统国际会议论文集，BuildSys ’22，第198–207页，美国纽约州纽约市，2022年a。计算机协会。ISBN 9781450398909。doi：10.1145/ 3563357.3564079。URL https://doi.org/10.1145/3563357.3564079。

Diptyaroop Maji、Ramesh K. Sitaraman 和 Prashant Shenoy。Dacf：使用机器学习预测电网日前碳强度。载于第十三届*ACM*未来能源系统国际会议论文集，e-Energy ’22，第188–192页，美国纽约州纽约市，2022年b。计算机协会。ISBN 9781450393973。doi：10.1145/3538637.3538849。URLhttps://doi.org/10.1145/3538637.3538849。

Gregory J Miller、Kevin Novan 和 Alan Jenn。电力消费碳排放的逐小时核算。*EnvironmentalResearchLetters*，17(4):044073，2022年4月。doi：10.1088/1748-9326/ac6147。URL https://dx.doi.org/10.1088/1748-9326/ac6147。

Fereydoun Farrahi Moghaddam。 面向高性能计算和 *Web* 应用的地理分布式云中碳利润感知型作业调度与负载均衡。博士论文，École de technologie sup´erieure，2014年。

英伟达。英伟达管理库（NVML），2022年。网址：https://developer.nvidia.com/nvidia-management-library-nvml。

David Patterson、Joseph Gonzalez、Quoc Le、Chen Liang、Lluis-Miquel Munguia、Daniel Rothchild、David So、Maud Texier 和 Jeff Dean。碳排放与大型神经网络训练，2021年。网址：https://arxiv.org/abs/2104.10350。

WattTime。WattTime，2022年。网址：https://www.watttime.org/。

Carole-Jean Wu、Ramya Raghavendra、Udit Gupta、Bilge Acun、Newsha Ardalani、Kiwan Maeng、Gloria Chang、Fiona Aga Behram、James Huang、Charles Bai、Michael Gschwind、Anurag Gupta、Myle Ott、Anastasia Melnikov、Salvatore Candido、David Brooks、Geeta Chauhan、Benjamin Lee、Hsien-Hsin S. Lee、Bugra Akyildiz、Maximilian Balandat、Joe Spisak、Ravi Jain、Mike Rabbat 和 Kim Hazelwood。可持续人工智能：环境影响、挑战与机遇，2021年。网址：https://arxiv.org/abs/2111.00364。

Jie Y ou、Jae-Won Chung 和 Mosharaf Chowdhury。Zeus：理解并优化 GP U en深度神经网络训练的能耗。发表于*USENIX NSDI*，2023 年。