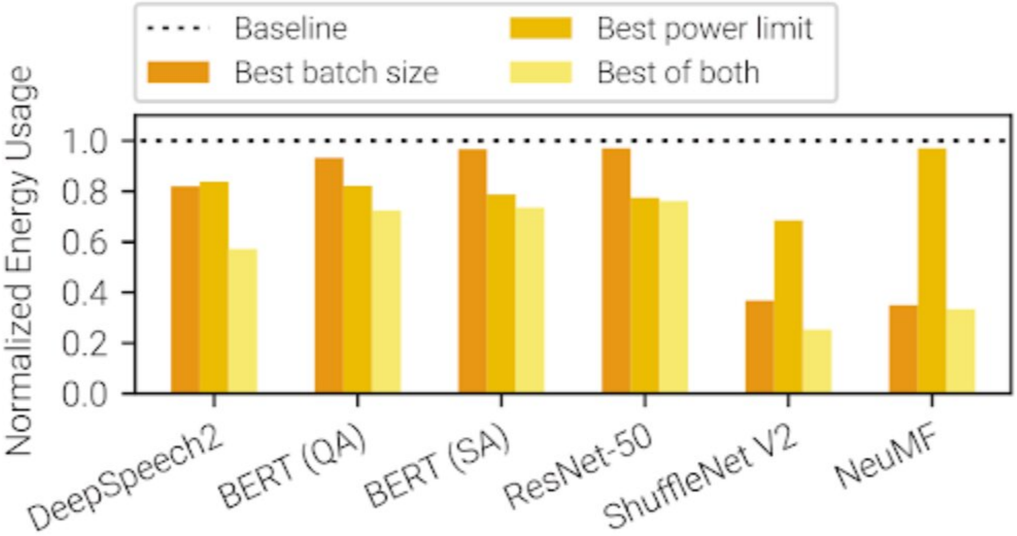
TechXplore

# Optimization could cut the carbon footprint of AI training by up to 75%

April 17 2023, by Kate McAlpine



Optimization could cut the carbon footprint of AI training by 15 to 75%. A variety of common deep learning models benefit from Zeus' ability to tune GPU power limits and the training batch size. When both parameters were tuned, the software achieved up to 75% energy reduction. Credit: SymbioticLab, University of Michigan

A new way to optimize the training of deep learning models, a rapidly evolving tool for powering artificial intelligence, could slash AI's energy demands.

Developed at the University of Michigan, the open-source optimization framework studies deep learning models during training, pinpointing the best tradeoff between energy consumption and the speed of the training.

"At extreme scales, training the GPT-3 model just once consumes 1,287 MWh, which is enough to supply an average U.S. household for 120 years," said Mosharaf Chowdhury, an associate professor of electrical engineering and computer science.

With Zeus, the new energy optimization framework developed by Chowdhury and his team, figures like this could be reduced by up to 75% without any new hardware—and with only minor impacts on the time it takes to train a model. It was presented at the 2023 USENIX Symposium on Networked Systems Design and Implementation (NSDI), in Boston.

Mainstream uses for hefty deep learning models have exploded over the past three years, ranging from image-generation models and expressive chatbots to the recommender systems powering TikTok and Amazon. With cloud computing already out-emitting commercial aviation, the increased climate burden from artificial intelligence is a significant concern.

"Existing work primarily focuses on optimizing deep learning training for faster completion, often without considering the impact on energy efficiency," said Jae-Won Chung, a doctoral student in computer science and engineering and co-first author of the study. "We discovered that the energy we're pouring into GPUs is giving diminishing returns, which allows us to reduce energy consumption significantly, with relatively little slowdown."

Deep learning is a family of techniques making use of multilayered, artificial neural networks to tackle a range of common machine learning tasks. These are also known as deep neural networks (DNNs). The models themselves are extremely complex, learning from some of the most massive data sets ever used in machine learning. Because of this, they benefit greatly from the multitasking capabilities of graphical processing units (GPUs), which burn through 70% of the power that goes into training one of these models.

Zeus uses two software knobs to reduce energy consumption. One is the GPU power limit, which lowers a GPU's power use while slowing down the model's training until the setting is adjusted again. The other is the deep learning model's batch size parameter, which controls how many samples from the training data the model works through before updating the way the model represents the relationships it finds in the data. Higher batch sizes reduce training time, but with increased energy consumption.

Zeus is able to tune each of these settings in real time, seeking the optimal tradeoff point at which energy usage is minimized with as little impact on training time as possible. In examples, the team was able to visually demonstrate this tradeoff point by showing every possible combination of these two parameters. While that level of thoroughness won't happen in practice with a particular training job, Zeus will take advantage of the repetitive nature of machine learning to come very close.

"Fortunately, companies train the same DNN over and over again on newer data, as often as every hour. We can learn about how the DNN behaves by observing across those recurrences," said Jie You, a recent doctoral graduate in computer science and engineering and co-lead author of the study.

Zeus is the first framework designed to plug into existing workflows for a variety of machine learning tasks and GPUs, reducing energy consumption without requiring any changes to a system's hardware or datacenter infrastructure.

In addition, the team has developed complementary software that they layer on top of Zeus to reduce the carbon footprint further. This software, called Chase, privileges speed when low-carbon energy is available, and chooses efficiency at the expense of speed during peak times, which are more likely to require ramping up carbon-intensive energy generation such as coal. Chase took second place at last year's CarbonHack hackathon and is to be presented May 4 at the International Conference on Learning Representations Workshop.

"It is not always possible to readily migrate DNN training jobs to other locations due to large dataset sizes or data regulations," said Zhenning Yang, a master's student in computer science and engineering. "Deferring training jobs to greener time frames may not be an option either, since DNNs must be trained with the most up-to-date data and quickly deployed to production to achieve the highest accuracy.

"Our aim is to design and implement solutions that do not conflict with these realistic constraints, while still reducing the carbon footprint of DNN training."

**More information:** Conference: www.usenix.org/conference/nsdi23

Study: Zeus: Understanding and Optimizing GPU Energy Consumption of DNN Training

Study: Chasing Low-Carbon Electricity for Practical and Sustainable DNN Training

Open-source software:

Zeus on GitHub

Chase on Github

Provided by University of Michigan

Citation: Optimization could cut the carbon footprint of AI training by up to 75% (2023, April 17) retrieved 26 April 2024 from https://techxplore.com/news/2023-04-optimization-carbon-footprint-ai.html