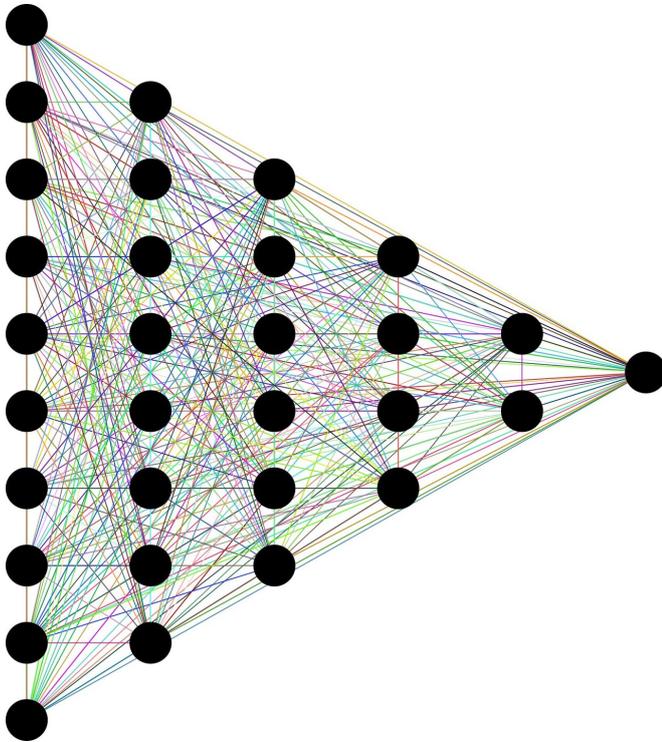# Accelerating the pace of machine learning

18 May 2022

Credit: CC0 Public Domain

Machine learning happens a lot like erosion.

Data is hurled at a mathematical model like grains of sand skittering across a rocky landscape. Some of those grains simply sail along with little or no impact. But some of them make their mark: testing, hardening, and ultimately reshaping the landscape according to inherent patterns and fluctuations that emerge over time.

Effective? Yes. Efficient? Not so much.

Rick Blum, the Robert W. Wieseman Professor of Electrical and Computer Engineering at Lehigh University, seeks to bring efficiency to distributed learning techniques emerging as crucial to modern artificial intelligence (AI) and machine learning (ML). In essence, his goal is to hurl far fewer grains of data without degrading the overall impact.

In the paper "Distributed Learning With Sparsified Gradient Differences," published in a special ML-focused issue of the *IEEE Journal of Selected Topics in Signal Processing*, Blum and collaborators propose the use of "Gradient Descent method with Sparsification and Error Correction," or GD-SEC, to improve the communications efficiency of machine learning conducted in a "worker-server" wireless architecture. The issue was published May 17, 2022.

"Problems in distributed optimization appear in various scenarios that typically rely on wireless communications," he says. "Latency, scalability, and privacy are fundamental challenges."

"Various distributed optimization algorithms have been developed to solve this problem," he continues," and one primary method is to employ classical GD in a worker-server architecture. In this environment, the central server updates the model's parameters after aggregating data received from all workers, and then broadcasts the updated parameters back to the workers. But the overall performance is limited by the fact that each worker must transmit all of its data all of the time. When training a deep neural network, this can be on the order of 200 MB from each worker device at each iteration. This communication step can easily become a significant bottleneck on overall performance, especially in federated learning and edge AI systems."

Through the use of GD-SEC, Blum explains, communication requirements are significantly reduced. The technique employs a data compression approach where each worker sets small magnitude gradient components to zero—the signal-processing equivalent of not sweating the small stuff. The worker then only transmits to the server the remaining non-zero components. In other words, meaningful, usable data are the only packets launched at the model.

"Current methods create a situation where each worker has expensive computational cost; GD-SEC

is relatively cheap where only one GD step is needed at each round," says Blum.

Professor Blum's collaborators on this project include his former student Yicheng Chen, PhD, now a [software engineer](#) with LinkedIn; Martin Taká?, an associate professor at the Mohamed bin Zayed University of Artificial Intelligence; and Brian M. Sadler, a Life Fellow of the IEEE, U.S. Army Senior Scientist for Intelligent Systems, and Fellow of the Army Research Laboratory.

More information: Yicheng Chen et al, Distributed Learning With Sparsified Gradient Differences, *IEEE Journal of Selected Topics in Signal Processing* (2022). [DOI: 10.1109/JSTSP.2022.3162989](#)

Provided by Lehigh University

APA citation: Accelerating the pace of machine learning (2022, May 18) retrieved 11 August 2022 from https://techxplore.com/news/2022-05-pace-machine.html