

Researchers show how network pruning can skew deep learning models

November 2 2022, by Matt Shipman



Credit: Pixabay/CC0 Public Domain

Computer science researchers have demonstrated that a widely used technique called neural network pruning can adversely affect the performance of deep learning models, detailed what causes these

performance problems, and demonstrated a technique for addressing the challenge.

Deep learning is a type of artificial intelligence that can be used to classify things, such as images, text or sound. For example, it can be used to identify individuals based on facial images. However, [deep learning models](#) often require a lot of computing resources to operate. This poses challenges when a [deep learning model](#) is put into practice for some applications.

To address these challenges, some systems engage in "[neural network pruning](#)." This effectively makes the deep learning model more compact and, therefore, able to operate while using fewer computing resources.

"However, our research shows that this network pruning can impair the ability of deep learning models to identify some groups," says Jung-Eun Kim, co-author of a paper on the work and an assistant professor of computer science at North Carolina State University.

"For example, if a security system uses deep learning to scan people's faces in order to determine whether they have access to a building, the deep learning model would have to be made compact so that it can operate efficiently. This may work fine most of the time, but the network pruning could also affect the deep learning model's ability to identify some faces."

In their new paper, the researchers lay out why network pruning can adversely affect the performance of the model at identifying certain groups—which the literature calls "[minority groups](#)"—and demonstrate a new technique for addressing these challenges.

Two factors explain how network pruning can impair the performance of deep learning models.

In technical terms, these two factors are: disparity in gradient norms across groups; and disparity in Hessian norms associated with inaccuracies of a group's data. In practical terms, this means that deep learning models can become less accurate in recognizing specific categories of images, sounds or text. Specifically, the network pruning can amplify accuracy deficiencies that already existed in the model.

For example, if a deep learning model is trained to recognize faces using a data set that includes the faces of 100 white people and 60 Asian people, it might be more accurate at recognizing white faces, but could still achieve adequate performance for recognizing Asian faces. After network pruning, the model is more likely to be unable to recognize some Asian faces.

"The deficiency may not have been noticeable in the original model, but because it's amplified by the network pruning, the deficiency may become noticeable," Kim says.

"To mitigate this problem, we've demonstrated an approach that uses mathematical techniques to equalize the groups that the deep learning model is using to categorize data samples," Kim says. "In other words, we are using algorithms to address the gap in accuracy across groups."

In testing, the researchers demonstrated that using their mitigation technique improved the fairness of a deep learning model that had undergone network pruning, essentially returning it to pre-pruning levels of accuracy.

"I think the most important aspect of this work is that we now have a more thorough understanding of exactly how network pruning can influence the performance of deep learning models to identify minority groups, both theoretically and empirically," Kim says. "We're also open to working with partners to identify unknown or overlooked impacts of

model reduction techniques, particularly in real-world applications for [deep learning](#) models."

The paper, "Pruning Has a Disparate Impact on Model Accuracy," will be presented at the 36th Conference on Neural Information Processing Systems (NeurIPS 2022), being held Nov. 28-Dec. 9 in New Orleans. First author of the paper is Cuong Tran of Syracuse University. The paper was co-authored by Ferdinando Fioretto of Syracuse, and by Rakshit Naidu of Carnegie Mellon University.

More information: "Pruning Has a Disparate Impact on Model Accuracy" Presented: Nov. 28-Dec. 9, 36th Conference on Neural Information Processing Systems (NeurIPS 2022)

Provided by North Carolina State University

Citation: Researchers show how network pruning can skew deep learning models (2022, November 2) retrieved 20 July 2024 from <https://techxplore.com/news/2022-11-network-pruning-skew-deep.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--