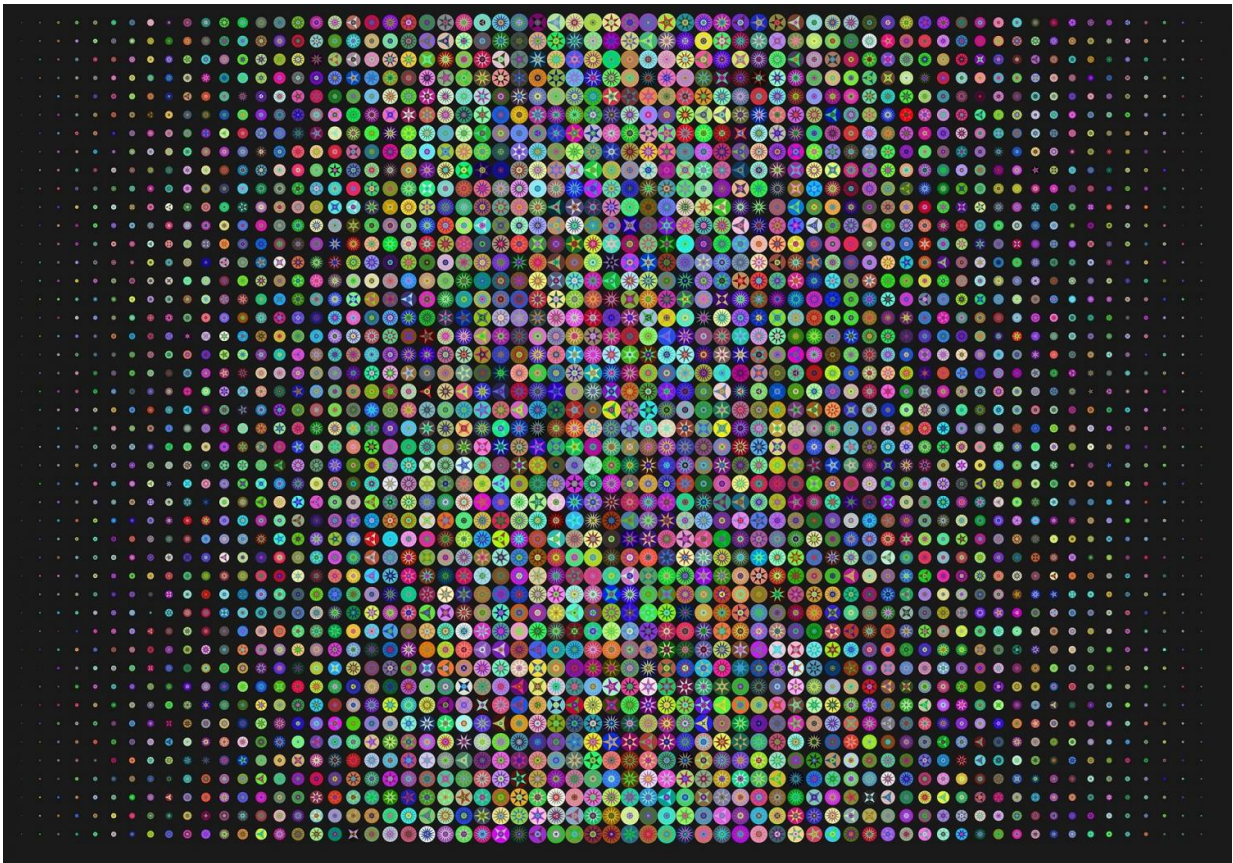


New algorithm limits bias in machine learning

July 16 2018, by Jenna Marshall



Credit: CC0 Public Domain

Machine learning—a form of artificial intelligence based on the idea that computers can learn from data and make determinations with little help from humans—has the potential to improve our lives in countless

ways. From self-driving cars to mammogram scans that can read themselves, machine learning is transforming modern life.

It's easy to assume that using algorithms for decision-making removes human [bias](#) from the equation. But researchers have found that [machine learning](#) can produce unfair determinations in certain contexts, such as hiring someone for a job. For example, if the data plugged into the [algorithm](#) suggest men are more productive than women, the machine is likely to "learn" that difference and favor male candidates over female ones, missing the bias of the input. And managers may fail to detect the machine's discrimination, thinking that an automated decision is an inherently neutral one, resulting in unfair hiring practices.

In a new paper published in the Proceedings of the 35th Conference on Machine Learning, SFI Postdoctoral Fellow Hajime Shimao and Junpei Komiyama, a research associate at the University of Tokyo, offer a way to ensure fairness in machine learning. They've devised an algorithm that imposes a fairness constraint that prevents bias.

"So say the credit card approval rate of black and white [customers] cannot differ more than 20 percent. With this kind of constraint, our algorithm can take that and give the best prediction of satisfying the constraint," Shimao says. "If you want the difference of 20 percent, tell that to our machine, and our machine can satisfy that constraint."

That ability to precisely calibrate the constraint allows companies to ensure they comply with federal non-discrimination laws, adds Komiyama. The team's algorithm "enables us to strictly control the level of fairness required in these legal contexts," Komiyama says.

Correcting for bias involves a trade off, though, Shimao and Komiyama note in the study. Because the constraint can affect how the machine reads other aspects of the data, it can sacrifice some of the machine's

predictive power.

Shimao says he would like to see businesses use the algorithm to help root out the hidden discrimination that may be lurking in their machine learning programs. "Our hope is that it's something that can be used so that machines can be prevented from discrimination whenever necessary."

Provided by Santa Fe Institute

Citation: New algorithm limits bias in machine learning (2018, July 16) retrieved 13 March 2024 from <https://phys.org/news/2018-07-algorithm-limits-bias-machine.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>
--