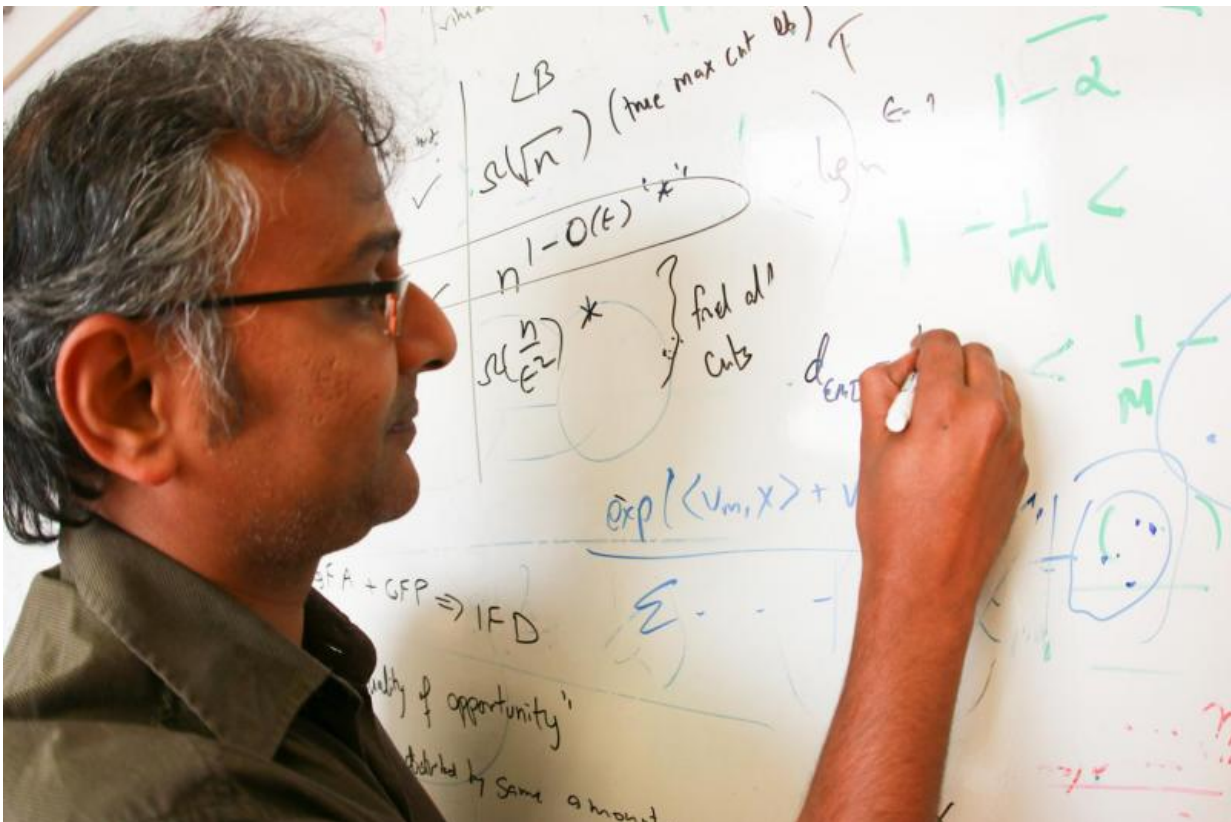


Programming and prejudice: Computer scientists discover how to find bias in algorithms

August 14 2015



Suresh Venkatasubramanian, an associate professor in the University of Utah's School of Computing, leads a team of researchers that have discovered a technique to determine if algorithms used for tasks such as hiring or administering housing loans could in fact discriminate unintentionally. The team also has discovered a way to fix such errors if they exist. Their findings were recently revealed at the 21st Association for Computing Machinery's SIGKDD

Conference on Knowledge Discovery and Data Mining in Sydney, Australia.
Credit: University of Utah College of Engineering

Software may appear to operate without bias because it strictly uses computer code to reach conclusions. That's why many companies use algorithms to help weed out job applicants when hiring for a new position.

But a team of computer scientists from the University of Utah, University of Arizona and Haverford College in Pennsylvania have discovered a way to find out if an algorithm used for hiring decisions, loan approvals and comparably weighty tasks could be biased like a human being.

The researchers, led by Suresh Venkatasubramanian, an associate professor in the University of Utah's School of Computing, have discovered a technique to determine if such software programs discriminate unintentionally and violate the legal standards for fair access to employment, housing and other opportunities. The team also has determined a method to fix these potentially troubled algorithms.

Venkatasubramanian presented his findings Aug. 12 at the 21st Association for Computing Machinery's Conference on Knowledge Discovery and Data Mining in Sydney, Australia.

"There's a growing industry around doing resume filtering and resume scanning to look for [job applicants](#), so there is definitely interest in this," says Venkatasubramanian. "If there are structural aspects of the testing process that would discriminate against one community just because of the nature of that community, that is unfair."

Machine-learning algorithms

Many companies have been using algorithms in [software programs](#) to help filter out job applicants in the hiring process, typically because it can be overwhelming to sort through the applications manually if many apply for the same job. A program can do that instead by scanning resumes and searching for keywords or numbers (such as school [grade point averages](#)) and then assigning an overall score to the applicant.

These programs also can learn as they analyze more data. Known as [machine-learning algorithms](#), they can change and adapt like humans so they can better predict outcomes. Amazon uses similar algorithms so they can learn the buying habits of customers or more accurately target ads, and Netflix uses them so they can learn the movie tastes of users when recommending new viewing choices.

But there has been a growing debate on whether machine-learning algorithms can introduce unintentional bias much like humans do.

"The irony is that the more we design artificial intelligence technology that successfully mimics humans, the more that A.I. is learning in a way that we do, with all of our biases and limitations," Venkatasubramanian says.

Disparate impact

Venkatasubramanian's research determines if these [software algorithms](#) can be biased through the legal definition of disparate impact, a theory in U.S. anti-discrimination law that says a policy may be considered discriminatory if it has an adverse impact on any group based on race, religion, gender, sexual orientation or other protected status.

Venkatasubramanian's research revealed that you can use a test to determine if the algorithm in question is possibly biased. If the test—which ironically uses another machine-learning algorithm—can accurately predict a person's race or gender based on the data being analyzed, even though race or gender is hidden from the data, then there is a potential problem for bias based on the definition of disparate impact.

"I'm not saying it's doing it, but I'm saying there is at least a potential for there to be a problem," Venkatasubramanian says.

If the test reveals a possible problem, Venkatasubramanian says it's easy to fix. All you have to do is redistribute the data that is being analyzed—say the information of the job applicants—so it will prevent the [algorithm](#) from seeing the information that can be used to create the bias.

"It would be ambitious and wonderful if what we did directly fed into better ways of doing hiring practices. But right now it's a proof of concept," Venkatasubramanian says.

Provided by University of Utah

Citation: Programming and prejudice: Computer scientists discover how to find bias in algorithms (2015, August 14) retrieved 2 May 2024 from <https://phys.org/news/2015-08-prejudice-scientists-bias-algorithms.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>
--