Computer scientists offer new techniques to measure social bias in software

August 17 2017, by Janet Lathrop

Today, banks are increasingly using software to decide who will get a loan, courts to judge who should be denied bail, and hospitals to choose treatments for patients. These uses of software make it critical that the software does not discriminate against groups or individuals, say computer science researchers at the University of Massachusetts Amherst.

Professor Alexandra Meliou in the College of Information and Computer Sciences says, "The increased role of software and the potential impact it has on people's lives makes software fairness a critical property. Data-driven software has the ability to shape human behavior: it affects the products we view and purchase, the news articles we read, the social interactions we engage in, and, ultimately, the opinions we form."

Meliou with professor Yuriy Brun and Ph.D. student Sainyam Galhotra, have developed a new technique they call "Themis," to automatically test software for discrimination. They hope Themis will empower stakeholders to better understand software behavior, judge when unwanted bias is present, and, ultimately improve the software.

Brun says, "Unchecked, biases in data and software run the risk of perpetuating biases in society. For example, prior work has demonstrated that racial bias exists in online advertising delivery systems, where online searches for traditionally-minority names were more likely to yield ads related to arrest records. Such software behavior can contribute to racial

stereotypes and other grave societal consequences."

The researchers' paper describing this research, published in pre-conference materials for the European Software Engineering Conference (ESEC/FSE 2017) before its September meeting in Paderborn, Germany, has won an Association of Computing Machinery Special Interest Group on Software Engineering (ACM SIGSOFT) Distinguished Paper Award. The work is supported by the National Science Foundation.

Brun explains that while earlier research has considered discrimination in software, Themis focuses on measuring causality in discrimination. Software testing allows Themis to perform hypothesis testing, to ask such questions as whether changing a person's race affects whether the software recommends giving that person a loan, he says.

"Our approach measures discrimination more accurately than prior work that focused on identifying differences in software output distributions, correlations or mutual information between inputs and outputs. Themis can identify bias in software whether that bias is intentional or unintentional, and can be applied to software that relies on machine learning, which can inject biases from data without the developers' knowledge," he adds.

When evaluated on public software systems from GitHub, Themis found that discrimination can sneak in even when the software is explicitly designed to be fair. State-of-the-art techniques for removing discrimination from algorithms fail in many situations, in part because prior definitions of discrimination failed to capture causality, the researchers point out.

For example, Themis found that a decision-tree-based machine learning approach specifically designed not to discriminate against gender was

actually discriminating more than 11 percent of the time. That is, more than 11 percent of the individuals saw the software output affected just by altering their gender.

Themis also found that designing the software to avoid discrimination against one attribute may increase discrimination against others. For example, the same decision-tree-based software trained not to discriminate on gender discriminated against race 38 percent of the time. "These systems learn discrimination from biased data, but without careful control for potential bias, software can magnify that bias even further," Galhotra says.

More information: Conference Paper: people.cs.umass.edu/~brun/pubs … bs/Galhotra17fse.pdf

Themis Project page: fairness.cs.umass.edu/

Provided by University of Massachusetts Amherst