Undergraduate Thesis Prospectus

Improving Algorithms of Criminal Risk Assessment Instruments

(technical research project in Computer Science)

Distrust of Algorithms in America's Criminal Justice System

(sociotechnical research project)

by

Rachel Choi

November 2, 2020

On my honor as a University student, I have neither given nor received unauthorized aid
on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.
Rachel Choi

Technical advisor: Nada Basit, Department of Computer Science

STS advisor: Peter Norton, Department of Engineering and Society

General Research Problem

Through the use of machine learning algorithms, how can criminal risk assessments be improved to produce more accurate decisions in America's criminal justice system?

The US imprisons more people than any other country in the world. As of 2020, the criminal justice system holds almost 2.3 million people in 1,833 state prisons, 110 federal prisons, 1,772 juvenile correctional facilities, 3,134 local jails, 218 immigration detention facilities, and 80 Indian County jails as well as in military prisons, civil commitment centers, state psychiatric hospitals, and prisons in the U.S. territories (Wagner, 2020). In other words, 1 in 38 adult Americans have been or are currently in under some form of correctional supervision. The overwhelmed criminal justice system has resulted in pressure to reduce prison numbers without risking a rise in crime. Thus, courtrooms across the US have turned to automated risk assessment tools in attempts to shuffle defendants through the legal system as efficiently and safely as possible (Hao, 2019). However, there have been ongoing debates over the tools used as critics have been raising concerns on potential racial bias in the algorithm.

Improving Algorithms of Criminal Risk Assessment Instruments

How can machine learning algorithms be improved in predicting a defendant's future risk for misconduct?

This research will be an independent project, which will be advised by Professor Nada Basit in the Computer Science Department. This project will investigate on current machine learning algorithms of the risk tools and identify features where bias can be introduced. This will help with improving the algorithms that are used in the current American justice system and potentially help reduce the racial tension raised with the tools. One of the constraints of the project would be obtaining actual algorithms of the risk assessment tools as the algorithms are produced by private companies, such as Northpointe, Inc. Thus, the goal of this project is to discover the underlying accuracy of the recidivism algorithms and to test whether the algorithm was biased against certain groups. With these results, I will identify features that might cause bias in the recidivism scores and suggest alternative algorithms to reduce potential bias.

Distrust of Algorithms in America's Criminal Justice System

In the U.S. since 2010, how have critics and defenders of criminal justice algorithms competed to influence the extent of their use?

Artificial Intelligence is widely used throughout the criminal justice system in the United States. The most commonly used are pretrial risk assessment algorithms, which are also called risk assessment tools, which are designed to predict a defendant's future risk for reoffending. They influence judgments about guilt or innocence, bail, and sentencing. However, the algorithms are largely hidden from public view, and critics contend they embed bias (Angwin et al., 2016).

Larson et al. (2016) found that COMPAS, a tool by Northpointe, was far more likely to incorrectly judge black defendants than white defendants to be at a higher risk of recidivism, and white defendants were more likely than black defendants to be incorrectly flagged as low risk. In fact, black defendants who were classified as a higher risk of violent recidivism did recidivate at a slightly higher rate than white defendants, and the likelihood ratio for white defendants was higher, 2.03, than for black defendants, 1.62 (Larson et al, 2016).

Participant groups include a panel of judges who support the automated system as they help judges rely on more than educated guesses in deciding what happens to the defendants. In fact, in a recent poll by the National Judicial College of 369 judges, a clear majority (65%) agreed that artificial intelligence can be a useful tool for combatting bias in bail and sentencing decisions, but it should never completely replace a judge's discretion (American Bar Association, 2020). Other participant groups are technologists and legal experts who are skeptical about the risk assessment algorithms. The technologists raise concerns on how machine-learning algorithms use statistics to find patterns in data. Thus, if historical crime data is fed to the algorithm, it will pick out the patterns associated with crime. However, these patterns are statistical correlations, not causations. Thus, the risk assessment algorithms will turn correlative insights into causal scoring mechanisms (Hao, 2019). Meanwhile, legal experts argue that the algorithms will make the legal system more incomprehensible and data-based, as courts have relied more on the automated tools when making decisions in the last few years (Re & Solow-Niederman, 2019). Participants also include community activists in American Civil Liberties Union (ACLU) and Black Lives Matter (BLM). ACLU activists argue that human prejudices can be baked into these tools because the machine-learning models are trained on biased police data (Larson & Schmidt, 2014). Similarly, activists in BLM argue that the risk assessment scores are measured with known sources of bias, such as "race" and "gang affiliation" (Sentencing Project, 2015). They argue that these features produce results that are biased towards their race.

References

American Bar Association. (2020, February 16). The good, bad and ugly of new risk-assessment tech in criminal justice. https://www.americanbar.org/news/abanews/aba-news-archives/2020/02/the-good--bad-and-ugly-of-new-risk-assessment-tech-in-criminal-j/

Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016). Machine Bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Hao, K. (2019, January 21). AI is sending people to jail-and getting it wrong. https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/

Heaven, W. (2020). Predictive policing algorithms are racist. They need to be dismantled. https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

Larson, E., & Schmidt, P. (Eds.). (2014). *The Law and Society Reader II*. NYU Press. JSTOR.

Larson, J., Mattu, S., Kirchner, L. & Angwin, J. (2016). How We Analyzed the COMPAS Recidivism Algorithm. Pro Publica. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Re, R. & Solow-Niederman, A. (2019). Developing Artificially Intelligent Justice. https://law.stanford.edu/wp-content/uploads/2019/08/Re-Solow-Niederman_20190808.pdf

Sentencing Project. (2015). Eliminating Racial Inequality in The Criminal Justice System. JSTOR.

Wagner, P. & Sawyer, W. (2020, March 24). Mass Incarceration: The Whole Pie 2020. https://www.prisonpolicy.org/reports/pie2020.html