

This document was produced based on notes taken during the Courts and Predictive Algorithms workshop of the Data & Civil Rights conference. This document represents a general summary of the discussion that took place. Not all attendees were involved in every part of the conversation, nor does this document necessarily reflect the views and beliefs of individual attendees. All workshop participants received workshop materials prior to the event to spark discussion. The primer can be found at:

http://www.datacivilrights.org/pubs/2015-1027/Courts_and_Predictive_Algorithms.pdf

Workshop Discussion Notes: Courts and Predictive Algorithms

Overview

This workshop gathered individuals drawn from law enforcement, social justice movements, technology companies, academia and government to discuss the effects of predictive algorithms on risk assessment in pre-trial and other stages of the judicial process. Throughout the session, there emerged significant differences of opinion over the underpinning values of the criminal justice system; if and how the system can be reformed was dependent on determining whether its flaws are fundamental to the way it currently operates. Along this axis of diverging viewpoints, some participants expressed strong support of algorithmic interventions in court settings, while others saw these technologies as challenges to justice and civil rights. The workshop had a number of goals, including an effort to draw out the great range of views present in the room. Organizers first focused on identifying and staking out the most controversial claims about the use of algorithms in courts. Using these prompts, the workshop allowed people to identify what they most valued, and which aspects of algorithmic interventions were most important to focus on in discussion. Ultimately, most participants did agree on several individual issues in need of change or reform: the use and setting of money bail, the potential for racial impact, the rate of incarceration, and the transparency of decision making. What remained contested among participants were the beliefs about which strategies would be most effective in improving the more problematic parts of court process. This begs the question: when will we have more data on these practices – a question that is complicated by the high-stakes process of judicial action.

Introduction

A range of algorithmic technologies and policies have begun to be used by judges and the associated juridical apparatus in recent years. These algorithmic tools can be used by different members of the court for different purposes, but the most visible and common cases involve judges receiving algorithmic suggestions for bail and sentencing decisions. These systems are often produced and maintained by a third party vendor, and the range of data used to produce predictive models is far from standardized. While there have been some efforts to make algorithmic court tools consistent, there are a number of

complicating factors, including unevenly distributed budgets and lobbyists from impacted industries (e.g. the Bail Bondsman).

With such a diverse set of techniques and practices which fall under the rubric of algorithmic courts, the workshop began by soliciting topics which participants felt were the most controversial positions to hold on the issue. The conversation was structured around asking participants to articulate their own positions in relation to the following statements:

“Money Bail Should Be Abolished”

The use of algorithms to set or make recommendations for bail determinations prompted a re-evaluation of the practice of money bail in any form. Supporters of the bail system felt that a move to eliminate the practice would be too far – “throwing the baby out with the bath water” – and suggested, instead, that the focus should be on reform. While early deployments may have been less than successful, algorithmic bail determination might be one method for reforming the system. Detractors of the money bail practice were not only skeptical of the reformative power of algorithms, but questioned the validity of the practice itself. They pointed to the asymmetrical racial impact of money bail, arguing that it is fundamental to the practice, and cannot be changed by the shifting of an algorithm’s threshold. Comparatively, there are many good models for this, such as Canada’s criminal justice system, which has no allowance for money bail.

“No Matter How Much I Distrust Algorithms, I Distrust Humans More”

The use of algorithms in court procedures is often compared to the use of human judgment in similar settings. Advocates of algorithmic methods (“trust the robots”) argue that human decision makers always use mental models to perform risk assessments and to make decisions. These mental models are simply more implicit than the explicit models used by algorithmic systems. Described in this way, humans are actually the *least* transparent algorithmic system, with internal biases invisible unless aggregated over large numbers. Proponents argued that various data bears this out, and judges have been shown to be risk-averse on average, meaning that they push to incarcerate when given the choice. However, the role algorithms might play in tempering this risk-aversion is unclear. Algorithmic systems might be able to formally incarnate rules that would prevent such “just to be safe” incarceration rates. Furthermore, the transparency of an algorithm is not a given, nor a constant – when and where is there information about which data the system uses, but also who developed the system and to what end goal?

“Risk-Assessment Techniques Need Years Before Implementation”

Some advocates of algorithmic systems challenged this idea. They argue that active or partial deployment of these techniques is the fastest and most efficacious way to develop, test, and improve them. Furthermore, this testing would be done within a judicial system that already has built-in procedures for

oversight, appeal, and redress. Cases based on faulty algorithmic data could be overturned at a later date. Those more skeptical of algorithmic systems were uncomfortable with this approach – citing the many possible unintended consequences of treating courts as experimental settings. They emphasized that the consequences of mistakes in this arena are dire, life-affecting ones. Existing data shows an increase in the incarceration rates of young black men, the population already dramatically over-impacted by the US criminal justice system.

“Are Racially Disproportionate Effects OK if the Intent was Racially Neutral?”

This position was met with unanimous disapproval from workshop participants. There is a clear parallel to draw here, as the establishment of mandatory minimum sentences was enacted with supposedly racially neutral goals, but has historically had racially discriminatory effects. In fact, the lingering impact of mandatory minimum sentences is one of the social ills that algorithmic systems might hope to correct. But while discriminatory effects were unanimously condemned, the specifics of how racially disparate effects are defined did not find a consensus in discussion. For example, should the results of algorithms be compared to the status quo, or should they be compared to existing crime data, or perhaps to other definitions of fairness and justice? In the algorithmic context, even without intent, racially disproportionate effects could result through the use of certain factors, such as geographies, which can serve as proxies for race. Finally, workshop participants also explored whether criminal justice systems might include “racial impact statements” modeled on the “environmental impact statements” that exist in industrial (and other) contexts. However, such an impact might be difficult to calculate, given the range of different policies that might be impacted by algorithmic systems.

Finding Common Ground

Following a period of discussion that brought out significant disagreements on many of the most controversial topics, participants set out to collect those values and assumptions that seemed to be shared by the group. There was broad agreement that the system of money bail was cause for concern (with or without algorithmic methods), that algorithms can have unintended impacts and can produce flawed risk assessments, as well as inspire too much confidence in their accuracy. The problems that algorithmic systems could help address are larger than just the judicial system (they are cultural and economic as well), but ultimately, any racially disparate impact is bad and should be corrected. Currently, there are huge gaps in knowledge about the data involved in algorithmic court procedures: these systems use bad and incomplete data, the data itself is often collected in problematic and inconsistent ways, and subsequent data about the efficacy of these systems is not adequately available. Participants agreed that judges and prosecutors should, in all cases, have some degree of discretion in decision making.

Participants also pooled their respective knowledge to create a sense of what the field of current algorithmic court techniques already looks like. Approximately 200 jurisdictions use some kind of pretrial algorithmic system, though there is no single generalized system. Most of the tools used are some variety of proprietary tool, and there are as of yet very few procedures for auditing the quality of these tools. The private and

proprietary nature of these tools is particularly problematic for critics, as the data is not available to assess their effectiveness or ethical soundness. An incomplete list of tools includes: SI-R (revised); Salient Factor Score; Static 99; and SOREG (sex offender recidivism guide).

Participants each described the jurisdictions and tools with which they were familiar, among which a set of themes emerged: that algorithmic tools should only be used to decrease (not increase) sentences, that tools should be audited in the case of vulnerable populations (especially youth), and that everywhere there was tussle over which data would be included in the system. Stakeholders in each case argued for the inclusion or non-inclusion of certain data points, such as location, employment, or age, based on individuals' values, goals, and experiences.

During this discussion, the more fundamental structure of algorithmic court procedure came to the fore. What, after all, was the goal of such algorithmic systems? Was it an instrument of civil rights? A tool for increased efficiency? Consistency? How is risk-assessment or recidivism defined? What is the difference between someone convicted 30 times of shoplifting and someone else convicted once of rape? How does a recidivism algorithm that is meant to perform risk-assessment handle such a case?

Conclusion

In the end, two points were emphasized as necessary steps to move forward in a positive and productive way. First, algorithmic stakeholders must acknowledge that the standard methods for improving algorithmic systems, such as testing and random trials, might not be available in the context of criminal justice, where a false positive, or inaccurate correlation could lead to real jail time for real individuals. The community, therefore, needs to derive productive ways of refining techniques that do not violate these ethical conditions. Secondly, all debate and development of algorithmic tools must proceed under the understanding that many of the intended users of these tools, including judges and prosecutors, are not and will not be proficient in statistics, data analysis, or issues of correlation and causation. Therefore, the on-the-ground use of these systems must be imagined in the hands of statistical non-experts, and programs to educate these individuals must be developed alongside the development of the algorithmic systems themselves.