Designing Equitable Risk Models for Lending and Beyond

1/7/21 6:01 PM

Speaker: Sharad Goel (Stanford Computational Policy Lab)

Overview

- Part 1: Many common mathematical definitions of algorithmic fairness are at odd with important understandings of equity
- Part 2: We can often design more equitable systems by explicitly separating prediction from decision making

Part I - Assessing Bias in Risk Models

- An example in criminal justice

Pretrial release decisions
- Shortly after arrest, judges must decide whether to release or detain defendants while they await trial
- Goal is to balance flight risk and public safety

Risk assessment tools:
- Assess the likelihood a defendant will fail to appear at trial or commit future crimes
- Risk is called risk of FTA (failure to appear)

Mathematical definition of fairness

- Classification parity:
  - Def1: Algorithm is fair if error rates are [approx] equal for while and black defendants
  - Def2: Algorithm has been tested, and the rate of error is balanced as between protected classes and those not in protected classes
- False positive rate
  - FPR = (Didn't reoffend & "high risk") / (Didn't reoffend)

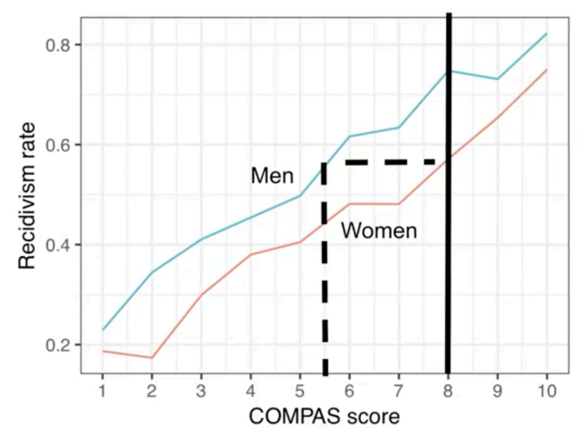
The problem of Infra-marginal

- FPR is an infra-marginal statistic - it depends not only on a group's threshold but on its distribution of risk
- Using the same threshold, higher risk groups will still have larger FPR, even if the No-Reoffend ratio in No-Release group is the same

Anti-classification

- Intuitively, a fair algorithm shouldn't use protected class

Problem with anti-classification

- Gender-neutral risk models can lead to discrimination
  - Woman are less likely to reoffend
  - 

  - COMPAS score: risk assessment; Recidivism rate: change of reoffend
  - If use a hard threshold, then we end up underestimating the risk of women, and overestimate the risk of men
- Can be fixed by using different models for different genders
- Or can be fixed by including gender in the model
  - Wisconsin uses gender-specific risk assessment tools

If data itself is biased, then there's no reason to trust the fairness of outcome

Measurement error

- Algorithm estimates the probability a defendant will be observed / reported committing a future violent crime
- Since reported crime is only a proxy for actual crime, estimates might be biased

Biased labels

- Algorithm tries to mimic past admissions decisions make by humans
- But past decisions were biased against women and minorities

Part II - Designing equitable algorithmic policies

Algorithms is not equal to policy

- Statistical algorithms are often good at synthesizing information to estimate risk. But we must still set equitable policy

Case Study: Inequities in lending

Motivation:
- 20% of U.S. households have no mainstream credit
- These households are disproportionately Black & Hispanic

How can we design a more inclusive lending policy?

Challenge:
- Allocate resources to underserved groups
- While maintaining relatively efficient (going loans to those who are more likely to repay)

Solution:
- Set different thresholds for different groups separately
- But it requires extra effort for screening (collecting information) for unbanked group

Selective screening

- Get more information on some individuals without mainstream credit who may in fact be creditworthy
  - E.g. reviewing household bills requires time and money
- Intuitively, we screen people "close" to the threshold, for whom the added information may plausibly make a difference in the lending decision

German credit experiment

1000 individuals, 70% are creditworthy

Two groups:

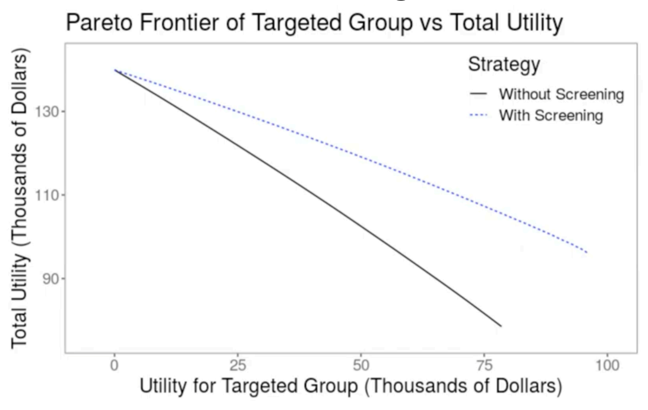
Those who own a residence [28%]

Those who don't [72%]

Greater proportion of homeowners are creditworthy [74% vs. 60%]

We assume the cost of screening is 10% the loan amount

Trade-off between utilities assigned to no-residence group vs. total utility



Summary

1. Equitable decision making generally requires examining the trade-off between competing concerns
   [Traditional fairness definitions are often overly rigid]
2. Important to understand the value of acquiring information and, more broadly, the value of interventions
   [Traditional fairness work treats information as static]

Reference

1. The measure and mismeasure of fairness: A Critical Review of Fair Machine Learning
2. Fair Allocation through Selective Information Acquisition