A Post-Processing Algorithm for Bias Reduction in Big Data Analytics
Roadmap

Motivation
- Why is this important?

Current Status
- Where are we at?

Future Directions
- What are our next steps?

Methodology
- Moment Multicalibration

Main Takeaways
- Summary
Machine Learning is becoming more prevalent but there are consequences...

**Why America Fails at Gathering Hate Crime Statistics**

The FBI relies on local law enforcement agencies to identify and report crimes, but their systems don't always capture hate crimes, even when they do occur.

**Artificial Intelligence**

How our data encodes systematic racism

Technologists must take responsibility for the toxic ideologies that our data sets and algorithms reflect.

**Facebook's ad-serving algorithm discriminates by gender and race**

Even if an advertiser is well-intentioned, the algorithm still prefers certain groups of people over others.

**In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation**

The bot learned language from people on Twitter—

**Racism and discrimination in health care: Providers and patients**

POSTED JANUARY 16, 2017, 9:30 AM, UPDATED JULY 09, 2020, 12:34 PM

Monique Tello, MD, MPH
Contributor

The Apple Card Didn't 'See' Gender—and That's the Problem

One way its algorithm determines credit lines makes the risk of bias more acute.
Let’s first ground our discussion...

**Key Observation**

The dosage prediction is averaged over the population, *not* an individual, so the dose might not be accurate for an individual.

- **Given features** $x$, your dosage for a drug is $f(x)$.
- **The variance conditional on my estimate** is $g(x)$.
- **For Asian Americans under the age of 50**, the confidence interval is $[a, b]$.
- **For women with a family history of diabetes**, the confidence interval is $[c, d]$. 

---

*How sure are you?*

*But I am part of a demographic representing less than 5% of the population*.
**Problem**

- Implicit bias against underrepresented populations in the systems we rely on
- Ethical standards of both fairness and privacy are breached

**Goal**

1) Make a tool that can supplement any existing algorithm, making it more fair
2) Reduce implicit bias
Solution:  **Multicalibration**

◎ **Calibration** assures that our predictions are accurate overall
  ○ Fails to make the same guarantee for subpopulations
  ○ E.g. 90% accuracy for the total population does not guarantee 90% accuracy for a subpopulation

◎ **Multicalibration** offers the same assurance across all possible subgroups
Project Components & Resources

Software
- Python
- Co
- PyTorch

Papers
- Moment Multicalibration for Uncertainty Estimation (Jung, Lee, Pai, Roth, Vohra)
- Multiaccuracy: BlackBox Post-Processing for Fairness in Classification (Kim, Ghorbani, Zou)

Stakeholders
- Cary Coglianese, Edward B. Shils Professor of Law and Professor of Political Science
- The Defender Association of Philadelphia
The Problem with ML in Criminal Justice

- **Data Description**: Data combines socio-economic data, law enforcement data, and crime data
- **Goal**: Predict violent crime number
- **Problem**: (1) High variance in accuracies for underrepresented people. (2) Models not calibrated to underrepresented people will only cause further harm
Algorithm Overview

**Auditor**
- Select subgroups
  - list of predefined subgroups
  - learning oracle algorithm
- Decide whether the subgroup prediction is calibrated

**Fixer**
- Adjust predictions for the chosen subgroup
- Return updated result to the Auditor

Repeat until Multicalibrated
Auditor Visualization

**Key Result**

This algorithm creates a classifier to predict points in a dataset that will likely have inaccurate predictions.

**residuals (n.)** The difference between the prediction and true label.
Fixer Visualization

Key Result
After $T$ iterations of post-processing, predicted mean is closer to the true mean.

Mean predictions adjusted during post-processing for validation data.

Mean predictions adjusted during post-processing for test data.
Main Takeaways

1. **Implemented** mean multicalibration based on the algorithm in the paper by Jung, Lee, Pai, Roth, Vohra
2. **Tested** the auditor and fixer on two different datasets
3. **Demonstrated** the results showing that both components work
Future Steps

1. Testing: *evaluate* the algorithm on a variety of datasets
2. Application: *run* algorithm on specific use cases such as housing and medication
3. Publishing: *put* our code on a publicly accessible site like IBM AI Fairness 360 Package for ML developers to utilize
Contact Us @
esesenior202021team06@gmail.com

Hyewon Lee

Trisha Pokharna

Brian Handen

Tashweena Heeramun

Margaret Ji