Fairness in Machine Learning

Solon Barocas

Cornell University

Fairness, Accountability, and Transparency in Machine Learning

Bringing together a growing community of researchers and practitioners concerned with fairness, accountability, and transparency in machine learning

The past few years have seen growing recognition that machine learning raises novel challenges for ensuring non-discrimination, due process, and understandability in decision-making. In particular, policymakers, regulators, and advocates have expressed fears about the potentially discriminatory impact of machine learning, with many calling for further technical research into the dangers of inadvertently encoding bias into automated decisions.

At the same time, there is increasing alarm that the complexity of machine learning may reduce the justification for consequential decisions to "the algorithm made me do it."

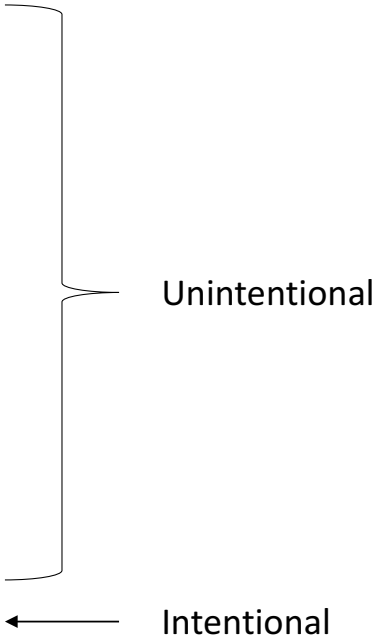
The annual event provides researchers with a venue to explore how to characterize and address these issues with computationally rigorous methods.

Many sources of bias

- Limited and coarse features
- Sample size disparity
  - Less data (by definition) about minority populations
- Skewed sample
  - Feedback loops
- Tainted examples
- Features that act as proxies
- Conscious prejudice

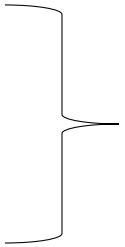
Many sources of bias

- Limited and coarse features
- Sample size disparity
  - Less data (by definition) about minority populations
- Skewed sample
  - Feedback loops
- Tainted examples
- Features that act as proxies
- Conscious prejudice

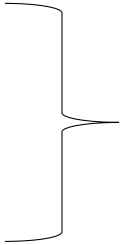


Many sources of bias

- Limited and coarse features
- Sample size disparity
  - Less data (by definition) about minority populations
- Skewed sample
  - Feedback loops
- Tainted examples
- Features that act as proxies
- Conscious prejudice



Detectable difference in accuracy rates for different parts of the population*



Undetectable difference

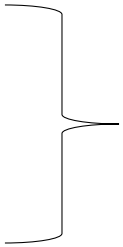
Depends how you view the problem

May be difficult for *outsiders* to detect

*If you know the protected class to which each person belongs

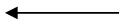
Addressing bias

- Limited and coarse features
- Sample size disparity
  - Less data (by definition) about minority populations
- Skewed sample
  - Feedback loops
- Tainted examples
- Features that act as proxies
- Conscious prejudice



Fairness by way of equal predictive accuracy; balance in type I and/or II error rates

Exploration to learn true base rates



Fairness by way of minimized disparity

Depends how you view the problem

What works for unintentional cases will work for intentional cases

Tensions

- Between different notions of fairness
- Between fairness and accuracy
- Between different methods for achieving fairness

Fairness, Accountability, and Transparency in Machine Learning

Bringing together a growing community of researchers and practitioners concerned with fairness, accountability, and transparency in machine learning

The past few years have seen growing recognition that machine learning raises novel challenges for ensuring non-discrimination, due process, and understandability in decision-making. In particular, policymakers, regulators, and advocates have expressed fears about the potentially discriminatory impact of machine learning, with many calling for further technical research into the dangers of inadvertently encoding bias into automated decisions.

At the same time, there is increasing alarm that the complexity of machine learning may reduce the justification for consequential decisions to "the algorithm made me do it."

The annual event provides researchers with a venue to explore how to characterize and address these issues with computationally rigorous methods.

Conference on Fairness, Accountability, and Transparency (FAT*)

A multi-disciplinary conference that brings together researchers and practitioners interested in fairness, accountability, and transparency in socio-technical systems.

The first FAT* conference will be held **February 23 and 24th, 2018 at New York University, NYC.**

The Call for Papers for FAT* 2018 is now available! Papers must be **registered by September 29**, and the **submission deadline is October 6, 2017**. The submission site is available here.

Algorithmic systems are being adopted in a growing number of contexts. Fueled by big data, these systems filter, sort, score, recommend, personalize, and otherwise shape human experiences of socio-technical systems. Although these systems bring myriad benefits, they also contain inherent risks, such as codifying and entrenching biases; reducing accountability and hindering due process; and increasing the information assymmetry between data producers and data holders.

FAT* is an annual conference dedicating to bringing together a diverse community to investigate and tackle issues in this emerging area. Topics of interest include, but are not limited to:

- The theory and practice of fair and interpretable Machine Learning, Information Retrieval, NLP, and Computer Vision
- Measurement and auditing of deployed systems
- Users' experience of algorithms, and design interventions to empower users
- The ethical, moral, social, and policy implications of big data and ubiquitous intelligent systems

FAT* builds upon several years of successful workshops on the topics of fairness, accountability, transparency, ethics, and interpretability in machine learning, recommender systems, the web, and other technical disciplines.

Announcements

- **October 4, 2017** Accepted papers at FAT* 2018 will appear in the Proceedings of Machine Learning Research.
- **September 22, 2017** The paper submission site for FAT* 2018 is now available.
- **August 31, 2017** Information about the FAT* 2018 conference is now available.
- **August 5, 2017** The FAT* website is now live. Details on the 2018 conference will be posted soon!