Perceptions of Fairness

Nripsuta Saxena, Karen Huang, Evan DeFilippis, Goran Radanovic, David C. Parkes, Yang Liu

USC Viterbi School of Engineering

UC SANTA CRUZ

The problem

- Multiple definitions (21! Maybe more!)
- No consensus
- Different situations, different definitions?
- Affects the public: their perceptions?

Project Motivation

- ProPublica and Northpointe (now Equivant) focus on different definitions of fairness
- Loans, bails, hiring, many more domains
- Impossibility results show some definitions cannot co-exist (Kleinberg et al., 2016)
- Which definitions appropriate for which contexts?

How might we understand people's perceptions of fairness in different contexts?

When is sensitive information ("protected attributes") important, and what (and how much) effect does it have?

Crowdsourcing

How do perceptions of fairness vary across geographies and cultures?

Example: loan decisions

- Divide $50,000 between two candidates
- Race does make a difference
- **So does gender!**
- Participants perceive race to be relevant
- Results suggest support for affirmative action
- US citizens: would results change in other cultures?

Extending further..

- Different cultures
- Different contexts, especially with indivisible goods: bail decisions, university admissions, hiring
- Indivisible goods
- To what extent these perceptions persist
- Distributive versus procedural justice
- All kinds of protected attributes treated similarly?
- *Why* does sensitive information matter?

Broad results

- Definitions in experiments vary in their strictness; People rated strictest definition to be most fair
- Sensitive information has an effect!
- People show more support to giving entire $50,000 to candidate with higher repayment rate, compared to splitting the money equally, when that candidate belongs to a historically disadvantaged group

How to incorporate public opinion?

- People may be directly affected by algorithmic decisions
- People can make inconsistent, unreasoned moral judgements (Greene, 2013)
- Moral machine show people approve of utilitarian autonomous vehicles, but unwilling to purchase utilitarian autonomous vehicles for themselves (Bonnefon et al., 2016)
- What to do when contradictory?
- *How* to blend the two together, and to what extent?

References

- Dwork, C., Hardt, M., Pitassi, T., Reingold, O. and Zemel, R., 2012, January. Fairness through awareness. In Proceedings of the 3rd innovations in theoretical computer science conference (pp. 214-226). ACM.
- Joseph, M., Kearns, M., Morgenstern, J.H. and Roth, A., 2016. Fairness in learning: Classic and contextual bandits. In Ad-vances in Neural Information Processing Systems (pp. 325-333).
- Liu, Y., Radanovic, G., Dimitrakakis, C., Mandal, D. and Parkes, D.C., 2017. Calibrated fairness in bandits. In Work-shop on Fairness, Accountability and Transparency in Ma-chine Learning, arXiv preprint arXiv:1707.01875.
- Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (ProPublica). Machine bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, 2016. Accessed: 2018-03-27.
- Dieterich, W., Mendoza, C. and Brennan, T. COMPAS risk scales: Demonstrating accuracy equity and predictive parity. http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf, 2016. Accessed: 2018-10-24.
- Bonnefon, J.F., Shariff, A. and Rahwan, I., 2016. The social dilemma of autonomous vehicles. Science, 352(6293), pp.1573-1576.
- Greene, J.D., 2013. Moral Tribes: Emotion, Reason, and the Gap between Us and Them. (Penguin).
- Emma Pierson. 2017. Gender differences in beliefs about algorithmic fairness. (2017). arXiv:1712.09124
- Nina Grgić-Hlača, Muhammad Bilal Zafar, Krishna Gummadi, and Adrian Weller. 2018. Beyond Distributive Fairness in Algorithmic Decision Making: Feature Selection for Procedurally Fair Learning. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI).
- Nina Grgić-Hlača, Elissa Redmiles, Krishna Gummadi, and Adrian Weller. 2018. Human Perceptions of Fairness in Algorithmic Decision Making: A Case Study of Criminal Risk Prediction. In Proceedings of the Web Conference (WWW).
- Narayanan, A. (2018). *21 fairness definitions and their politics*. Conference on Fairness, Accountability, and Transparency, February 23, New York.