Algorithmic Fairness and Algorithmic Explainability in Finance

Doaa Abu Elyounes

* *Harvard Law School - Berkman Klein Center for Internet and Society*
* *École Normale Supérieure, Paris*

dabuelyounes@sjd.law.harvard.edu

Related Paper:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3478296

Artificial Intelligence in Consumer Finance: Defining and Insuring Fairness
Conference, Online, 9 November 2021

Introduction

- Accuracy versus fairness.
- Explainable algorithms are perceived to be less accurate but more fair.
- Unexplainable algorithms are perceived to be more accurate but less fair.
- The degree of interpretability should be determined based on the domain the algorithm is implemented in.

Introduction (contd.)

- Researchers focusing on explainable algorithms are working on methods for increasing the accuracy rate of interpretable algorithms.
- Researchers focusing on unexplainable algorithms are trying to introduce more fairness into algorithms.
- Three categories of algorithmic fairness: individual fairness, group fairness, and causal reasoning.

- Narayanan, A. (2018) **21 Fairness Definitions and their Politics**, *The Conference on Fairness, Accountability and Transparency in Machine Learning FAT*2018*.
- Verma, S. and Rubin, J. (2018) **Fairness Definitions Explained**, *Proceedings of the International Workshop on Software Fairness.*

Individual Fairness and the Case of Color Blindness: The Unaware Approach

- Equal protection/ colorblindness
- 14th Amendment to the U.S. Constitution- "[…] no State shall […] deny to any person within its jurisdiction the equal protection of the laws."
- From the legal and computational perspective, colorblindness did not work

In which cases individual fairness will still work?

Individual Fairness (contd.)

Fairness Through Awareness

- Similarly situated individuals should be treated similarly
- Defining the mathematical metric
- Regulation by its nature seeks to differentiate.

- Dwork, C. et al., (2012), **Fairness Through Awareness**, Proceedings of the 3rd Innovations in Theoretical Computer Science Conference

Group Fairness and The Case of Affirmative Action

- Group fairness approaches acknowledge the circumstances that lead different groups to react differently to a given situation
- Protected attributes are addressed in the equation
- The focus is on the outcome
- Legally group fairness = affirmative action

1) Decoupling

- One algorithm per group
- The list of predictors may vary across groups
- COMPAS Men and COMPAS Women
- Is it socially acceptable to use a different algorithm for classifying minorities?

2) Statistical Parity

- The fraction of people from group A who receive a particular outcome is the same as the fraction of group A of the whole population
- Does not take into account the difference in the base rate across groups
- In which cases statistical parity can still be useful?

3) Conditional Statistical Parity

- Specific case of statistical parity
- Calls for equalizing each one of the factors

4) Equal Opportunity

- Individuals who qualify for a desirable outcome should have an equal chance of being correctly classified for this outcome
- Equalizing the true positives
- Doesn’t take into account the disparities among those who will be classified as high risk

5) Equalized Odds/ Equal Accuracy

- Equalizing the errors that the algorithm make across groups
- Equalizing false positives and false negatives
- Hard to achieve because of the following reasons ….

Balancing between False Negatives and False Positives

- What is the error rate that our society is willing to tolerate?
- The utilitarian approach- protecting financial institutions from failure, trust in financial institutions
- The egalitarian/ individual justice approach- everyone should get equal access to credit

Calibration

- Probabilities should carry semantic meaning
- An algorithm that has been calibrated is an algorithm that achieved equality within any given score category that it creates
- Calibration is important for maintain trust in the algorithm

Causal Reasoning and Due Process

- Correlation does not imply causation
- The black box problem and lack of explainability
- Risk of jeopardizing due process
- Causal Reasoning Approaches- only factors that directly cause the outcome will be included in the model
- Counterfactual fairness

We Cannot Satisfy all Notions of Fairness Simultaneously

- Chouldechova, A. (2017) **Fair Prediction,** arXiv: 1610.07524
- Friedler, S. Scheidegger C. Venkatasubramanian, S. Choudhary, S. Hamilton, E.P. and Roth, D. (2018) **A Comparative Study of Fairness- Enhancing Interventions in Machine Learning**, arXiv: 1802.04422

What Developers Can Do?

- Clarifying their approach to fairness
- Increased social and cultural understanding

What Policy Makers Can Do?

- Clarifying the laws and regulations
- Auditing
- Encouraging interdisciplinarity

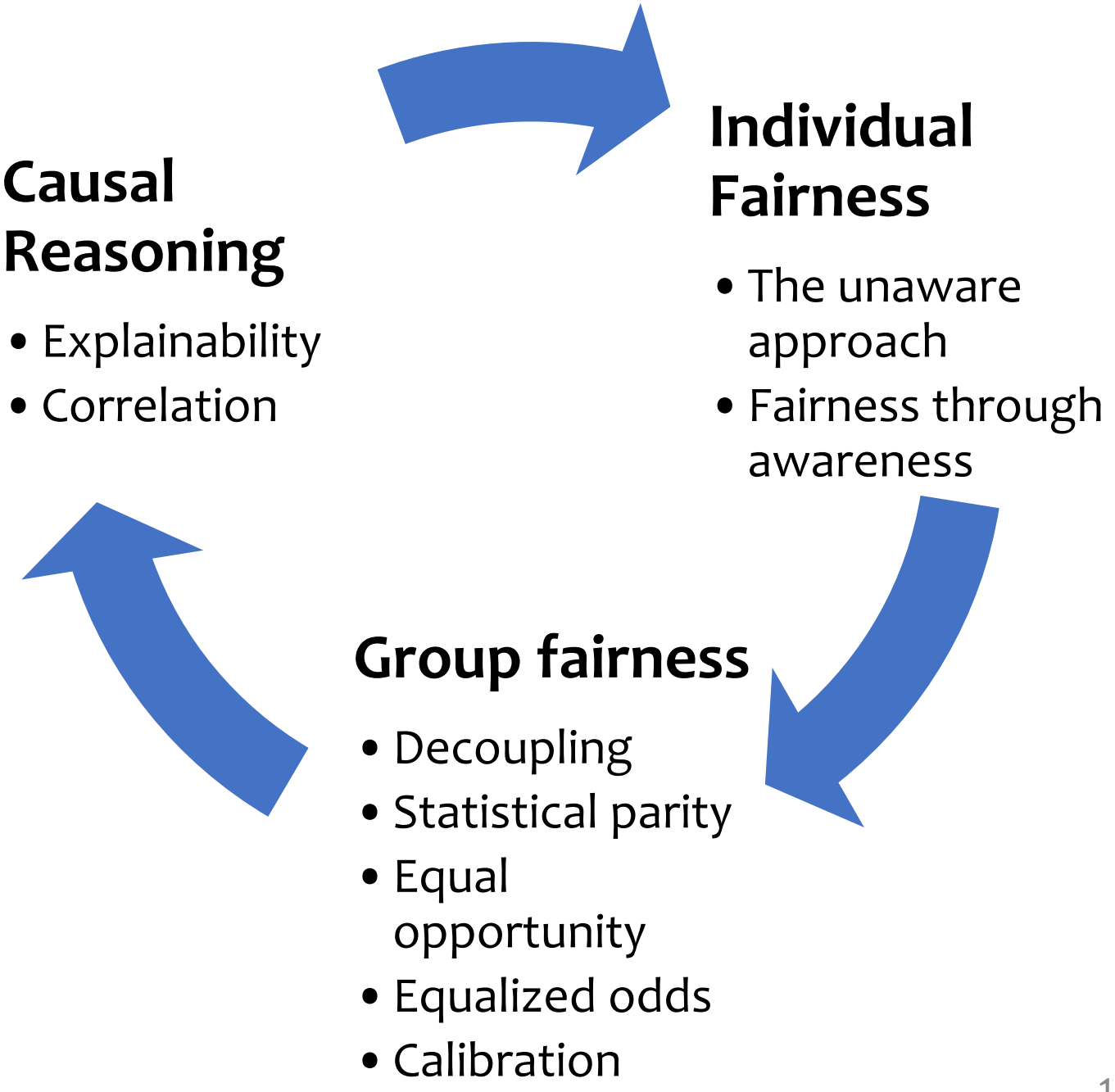
Conclusions

~~"All models are wrong, but some are useful"~~

~~*George Box, The Statistician*~~

Most models are right, but it depends how we use them

Conclusions:



Thank you

dabuelyounes@sjd.law.harvard.edu

For more details:

Doaa Abu Elyounes, "**Contextual Fairness: A Legal and Policy Analysis of Algorithmic Fairness**", University of Illinois Journal of Law, Technology & Policy (JLTP), 2020