Optimization under Fairness Constraints

Project Proposal for

Causal Inference

Jing Gong

November 1, 2017

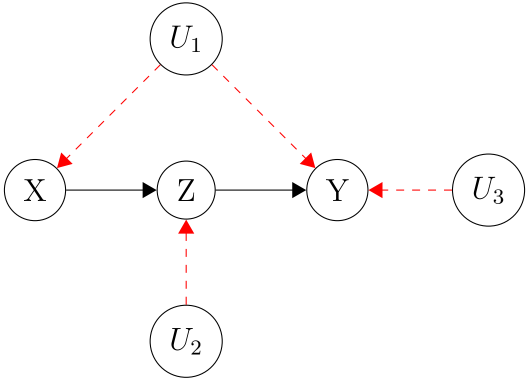
1 Introduction

Algorithms from statistics and machine learning are increasingly being used to make decisions that play important roles on people's lives, including policing, pricing, hiring, criminal sentencing and so on. Many of the decision making process apply machine learning techniques and apply historical data to make decisions. However, if the historical data contains discrimination, the machine learning models are very likely to learn the relationship of discrimination and reveal it in new decisions. This happens when sensitive attributes information has disparate impacts to the outcomes. Such sensitive attributes usually include gender, religion, physical ability and so on. These are referred as the protected attributes in Federal Laws and regulations. Take criminal sentencing as an example, blacks were more than twice as likely as whites to be labeled risky even though they didn't continue to commit a crime.

Hence, it is a curial concern in decision making process to put fairness into consideration. That is to day, decision process should be fair and decision results should be balanced. One thing should be noticed is the causal effect. This project aims to explore a novel causal approach to balance the decision results based on fairness.

2 Problem Setup

1. A Causal structural model like this:



2. Define causal fairness criterion: the causal effect between sensitive attributes and decisions should be the same for different values of sensitive attributes.

3. Optimization can be done under this fairness constraints

3 Related Work

The research work on exploring fairness algorithms is interdisciplinary. Different applications and different ways of fairness measurements are applied rigorously. Without changing the original training process, [3] derives an optimal equalized odds threshold predictor via a post-processing step. There are other works that develop new algorithms to measure fairness. For example, [2] designs a decision rule in the criminal justice system, focusing on the balance of benefit of releasing defendants and economic costs of detention. Policymaker would prefer the one that maximizes the utility when statistical parity is guaranteed. [4] introduces a measurement of decision boundary in the mechanism of fair classifiers, which applies decision boundary covariance to sensitive attributes. This paper gives two formulations: one that ensures a non-discrimination policy; one that ensures certain business necessity clause.

Compared to other literatures, [5] introduces a causal framework to remove direct and indirect discrimination. This paper gives an idea that measuring the causal effect of all connections removes the unfairness of protected attributes from data.

4 Reference

[1] M. Joseph, M. Kearns, J. Morgenstern, A. Roth. Fairness in Learning: Classic and Contextual Bandits, 2016.

[2] S. C. Davies, E. Pierson, A. Feller. Algorithmic Decision Making and the Cost of Fairness, 2017

[3] M. Hardt, E. Price, N. Srebro. Equality of Opportunity in Supervised Learning, 2016

[4] M. B. Zafar, I. Valera, M. G. Rodriguez, K. P. Gummadi. Fairness Constraints: Mechanisms for Fair Classification, 2017

[5] L. Zhang, Y. Wu and X, Wu. A Causal Framework for Discovering and Removing Direct and Indirect Discrimination, 2016

[6] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias: There?s software used across the country to predict future criminals. and it?s biased against blacks. ProPublica (5 2016).