Fairness in Machine Learning SS21

Beyond Static Fairness

Daniel Saggau

03/05/2021

Dynamical Repercussions

- Apple card discriminating women

Introduction

- Fairness in dynamical system has become important because equalizing true positive rate at each step does not converge as fast in systems with e.g. population dynamics
- Research on dynamical systems has focused on markov decision processes
- Prior research on causal fairness has focused in static systems

Fairness-Gym

- Environment to simulate fairness

Why care about causal fairness models?

- Fairness is not static (D'Amour et al. 2020) introduce shortcomings of existing fairness correction measures
- From a modelling perspective, we can improve the specification (Markov Decision Process)
- Regular models may express a model in conditional probabilities (probabilistic model) or may be expressed as differential equation
- SCMs are cast in functional form which is more stable (but may also be expressed as differential equation)
- Additionally, we specify the model beyond conditional probabilities -> we specify the latent variables (exogenous variables not observable within our dataset)

What is a SCM

- Functional specification of our model including latent factors
- Probabilistic specification + additional knowledge

SCMs for Fairness in dynamical Systems

- Fair-MDP
- Judea Pearl:do-calculus

Graphical Illustration

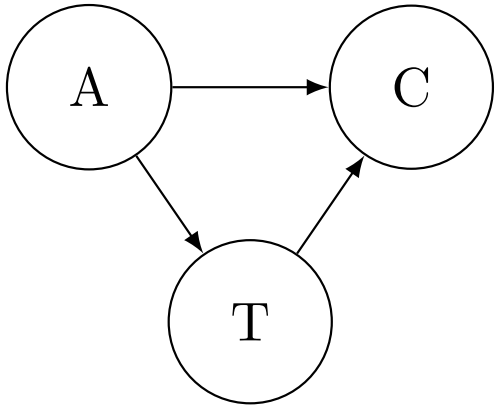

Figure 1: Probabilistic Model

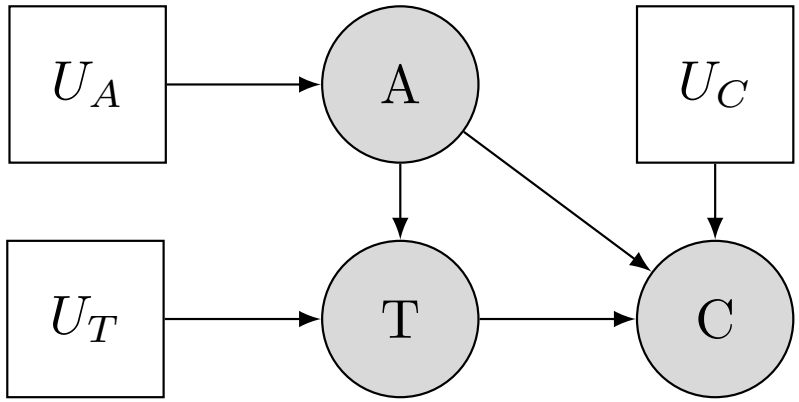

Figure 2: Structural Causal Model

Interventions

- Atomic Intervention
- Policy Intervention
- Off-Policy Intervention (model-based; model-free)

Results Off Policy Intervention - Lambda

- Introduce trade off parameter lambda
- V of Pi is the overall objective, Pi is our policy and U is utility

$$V_\pi = U - \lambda \delta_{EQOPP}$$

Causal Hierachy

Table 1: Pearls Hierachy of Causation (2009)

Method	Action	Example	Usage
Association $P(a\|b)$	Co-occurrence	What happened...	(Un-)Supervised ML, BN, Reg.
Intervention $P(a\|do(b), c)$	Do-manipulation	What happens if ...	CBN,MDP,RL
Counterfactual $P(a_b\|a', b')$	Hypotheticals	What would have happened if...	SCM ,PO

Exisiting Considerations

- Counterfactuals for Fairness in Dynamical Systems
- Off-policy estimation (model based (regression) or model free estimation (propensity weight))

Adding to existing Literature

Methodological:

- Cyclic Structural Causal Models with actual reinforcing loops
- Semi-Deterministic SCMs (deterministic -> all variables are known)
-

Application:

- Off Policy Interventions