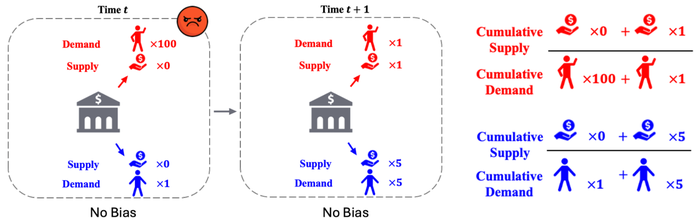
Adapting Static Fairness to Sequential Decision-Making: Bias Mitigation Strategies towards Equal Long-term Benefit Rate

Yuancheng Xu* [1] (ycxu@umd.edu), **Chenghao Deng*** [1], **Yanchao Sun**[1], **Ruijie Zheng**[1], **Xiyao Wang**[1], **Jieyu Zhao**[2], **Furong Huang**[1]

[1] University of Maryland, College Park [2] University of Southern California

Introduction

❖ Long-term fairness: A bank lending example
- *Demand*: number of applicants; *Supply*: number of loans



Both acceptance rates = 0 (Fair at Time t)

Both acceptance rates = 1 (Fair at Time t+1)

Not fair in the long run!

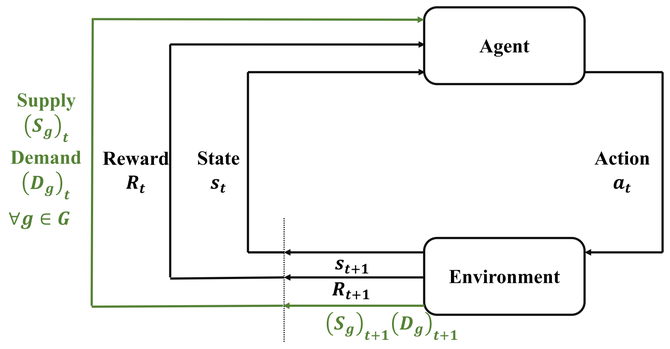
❖ Future impact matters

❖ Previous long-term fairness notions: summing up step-wise biases
- Temporal discrimination within groups
- Short-term fairness implies long-term fairness (Incorrect!)

Contributions
- Long-term fairness notion: Equal Long-term Benefit Rate (**ELBERT**)
  - Adapts multiple static fairness notions (e.g., Equal Opportunity)
  - A framework in Markov Decision Process (MDP)
- Bias mitigation algorithm using policy optimization

Equal Long-term Benefit Rate (ELBERT)

❖ Supply-Demand Markov Decision Process (SD-MDP)
- Additionally returns group demand and group supply as fairness signals



- Cumulative group supply and cumulative group demand

$$\eta_g^S(\pi) := \mathbb{E}_\pi\left[\sum_{t=0}^{\infty}\gamma^t S_g(s_t,a_t)\right], \eta_g^D(\pi) := \mathbb{E}_\pi\left[\sum_{t=0}^{\infty}\gamma^t D_g(s_t,a_t)\right]$$

❖ (Definition) Long-term Benefit Rate: $\frac{\eta_g^S(\pi)}{\eta_g^D(\pi)}$
- Ratio between cumulative group supply and demand

❖ (Definition) Bias of a policy

$$b(\pi) = \max_{g\in G}\frac{\eta_g^S(\pi)}{\eta_g^D(\pi)} - \min_{g\in G}\frac{\eta_g^S(\pi)}{\eta_g^D(\pi)}$$

- Discrepancy of Long-term Benefit Rate among groups

Bias Mitigation Algorithm

❖ Training objective (two groups)

$$\max J(\pi) = \eta(\pi) - \alpha b(\pi)^2 = \eta(\pi) - \alpha\left(\frac{\eta_1^S(\pi)}{\eta_1^D(\pi)} - \frac{\eta_2^S(\pi)}{\eta_2^D(\pi)}\right)^2$$

- Increase reward & decrease bias
- Challenge: policy gradient of Long-term Benefit Rate $\frac{\eta_g^S(\pi)}{\eta_g^D(\pi)}$

❖ Policy Optimization (PO)
- Key: reduction to standard policy gradient

$$\nabla_\pi\left(\frac{\eta_g^S(\pi)}{\eta_g^D(\pi)}\right) = \frac{1}{\eta_g^D(\pi)}\nabla_\pi\eta_g^S(\pi) - \frac{\eta_g^S(\pi)}{\eta_g^D(\pi)^2}\nabla_\pi\eta_g^D(\pi)$$

- Advantage function

$$\nabla_\pi J(\pi) = \mathbb{E}_\pi\{\nabla_\pi \log\pi(a_t|s_t)A_t^{fair}\}$$

- Can be plugged into any PO methods, like PPO
- Result

$$b(\pi)^2 = h\left(\frac{\eta_1^S(\pi)}{\eta_1^D(\pi)}, \frac{\eta_2^S(\pi)}{\eta_2^D(\pi)}\right)$$

$$A_t^{\text{fair}} = A_t - \alpha\sum_{g\in G}\frac{\partial h}{\partial z_g}\left(\frac{1}{\eta_g^D(\pi)}A_{g,t}^S - \frac{\eta_g^S(\pi)}{\eta_g^D(\pi)^2}A_{g,t}^D\right)$$

❖ Extension to multi-group setting
- Challenge: non-smoothness in max/min

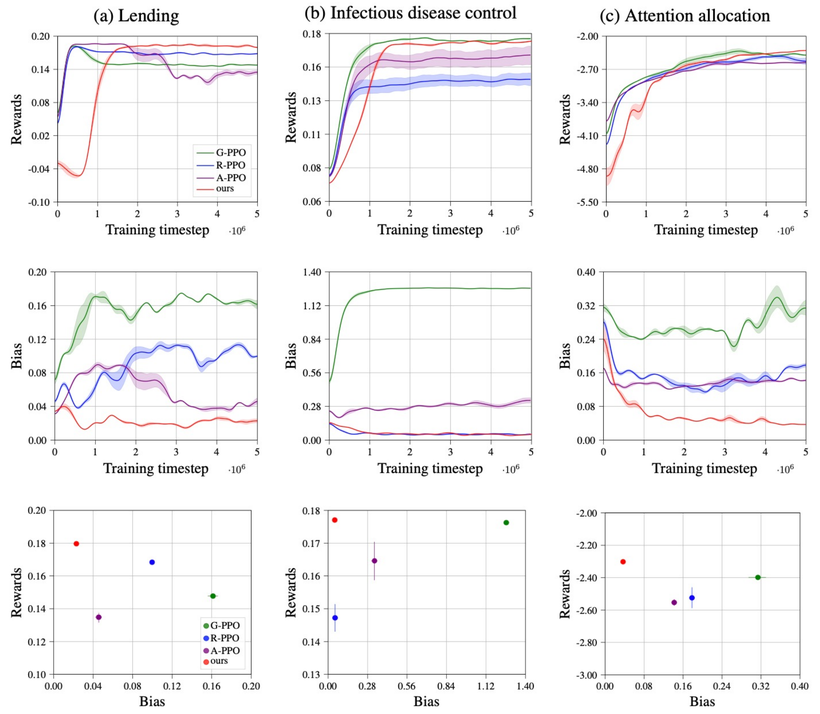
$$J(\pi) = \eta(\pi) - \alpha b(\pi)^2 = \eta(\pi) - \alpha\left(\max_{g\in\mathcal{G}}\frac{\eta_g^S(\pi)}{\eta_g^D(\pi)} - \min_{g\in\mathcal{G}}\frac{\eta_g^S(\pi)}{\eta_g^D(\pi)}\right)^2$$

- Solution: define soft bias

$$b^{\text{soft}}(\pi) = \frac{1}{\beta}\log\sum_{g\in G}\exp\left(\beta\frac{\eta_g^S(\pi)}{\eta_g^D(\pi)}\right) - \frac{1}{-\beta}\log\sum_{g\in G}\exp\left(-\beta\frac{\eta_g^S(\pi)}{\eta_g^D(\pi)}\right)$$

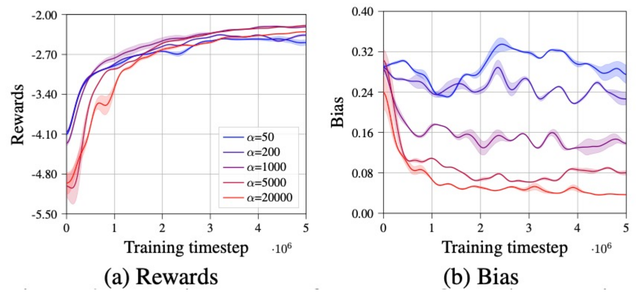
Experiments

❖ **ELBERT**-PO signifincantly reduces bias while maintaining high reward



Rewards and bias of **ELBERT**-PO and baselines in three environments

❖ Effect of the bias coefficient α



(a) Rewards (b) Bias

Larger the bias coefficient α
- Lower bias
- Slower convergence of reward. Final reward might be higher.