Long-Term Fair Decision Making through Deep Generative Models

Yaowei Hu[1], Yongkai Wu[2], Lu Zhang[1]
[1]University of Arkansas, [2]Clemson University

INTRODUCTION

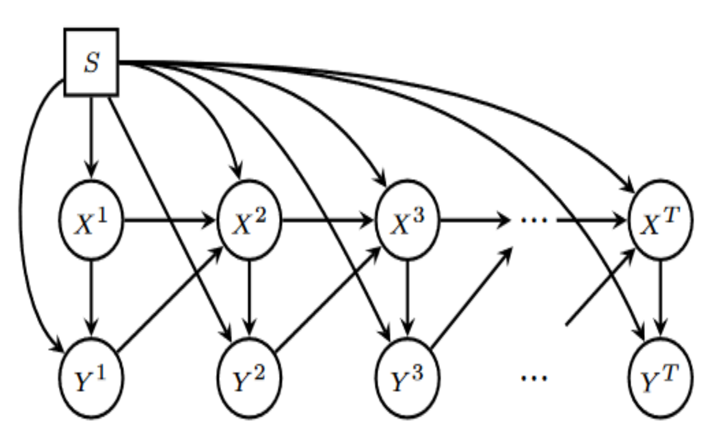
- The US Supreme Court ruled that affirmative action policies are unconstitutional and race can no longer be regarded as a factor in admissions to US universities. It shows group disparities and fairness are important aspects of social concern.
- Currently, the majority of studies in fair machine learning are focused on the problem of building decision models for fair one-shot decision-making. However, the algorithms based on traditional fairness notions cannot mitigate group disparities and could even exacerbate the gap.
- Long-term fairness has been proposed to focus on the mitigation of group disparities in the sequential decisions rather than making fair decisions in a single time step.

Our Goal

- We mitigate group disparity and achieve long-term fairness while limiting the use of the sensitive attribute in decision-making models.

PROBLEM SETTING

- Long-term Fairness for Sequential Decision Making
  - Given a time time series dataset $\mathcal{D}=\{(S,\mathbf{X}^t,Y^t)\}_{t=1}^{l}$ and causal graph.
  - S is a binary sensitive feature.
  - $\mathbf{X}^t$ is the profile features at time step t.
  - Y^t is a binary decision based on S and $\mathbf{X}^t$.



Our Task

- Learn a decision model $h_\theta : \mathcal{S}\times\mathcal{X}\mapsto\mathcal{Y}$ such that when deployed at every time step, fairness can be achieved at a certain time step T where $T > l$.

Analysis from the Causality Perspective

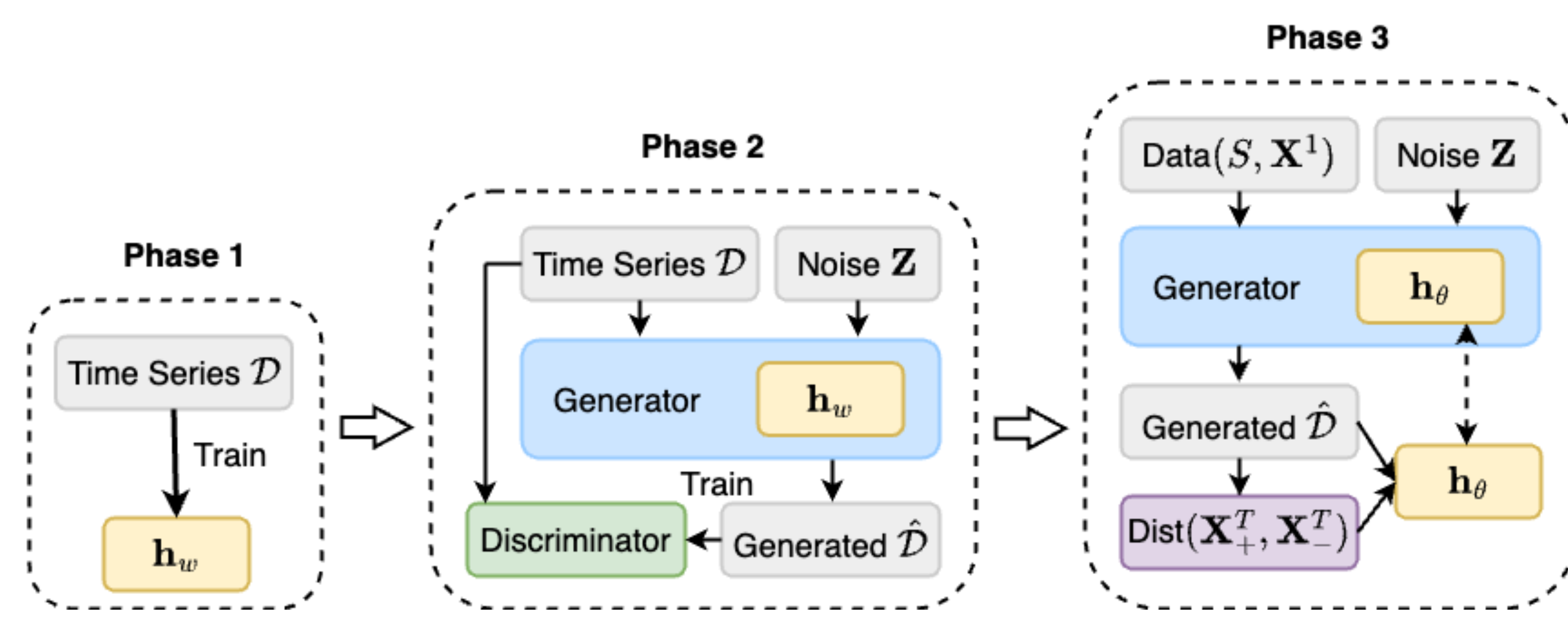
- Paths of causal effect of S on $\boldsymbol{X}^T$ can be categorized into:
  - Those that intersect with decision nodes (e.g., $S\rightarrow \boldsymbol{X}^1\rightarrow Y^1\rightarrow \boldsymbol{X}^2\rightarrow$ ...).
  - Those that bypass decision nodes (e.g., $S\rightarrow \boldsymbol{X}^1\rightarrow \boldsymbol{X}^2\rightarrow$ ...).
- Eliminating the causal effect of S on $\boldsymbol{X}^T$ via updating the decision model means learning a decision model h_θ such that the causal effects transmitted through two sets of paths are cancelled out.
- Due to the requirement of sensitive attribute unconsciousness, long-term fairness may not always be achievable on through updating the decision model.

METHODOLOGY

Core Ideas

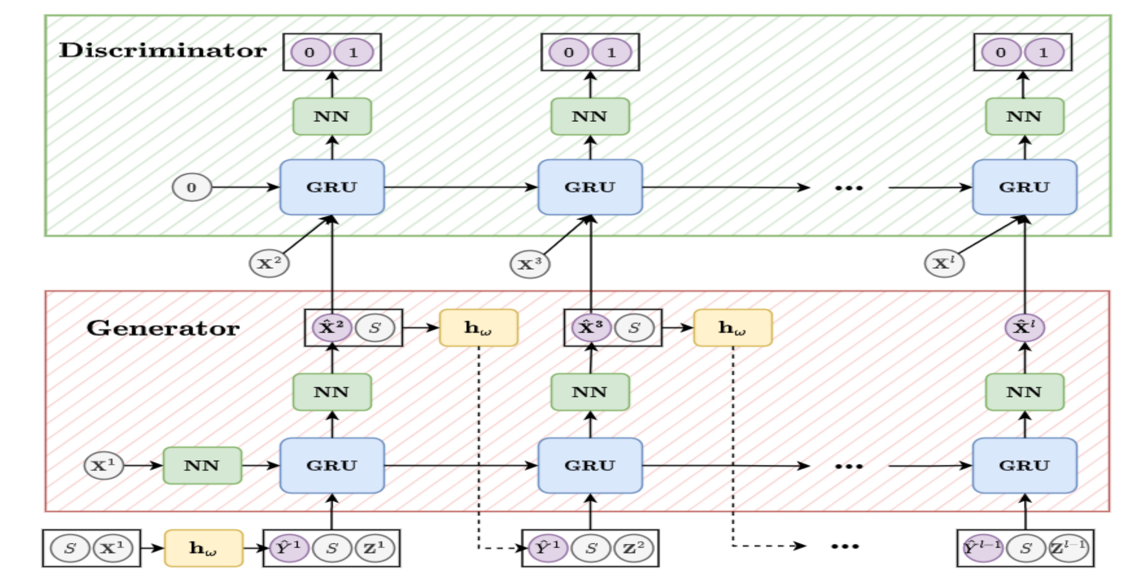
- Design a deep generative model predictively generate data following both observational and interventional distributions.
- Integrate the prediction and model training into a collaborative training framework such that the predicted data are used as reliable data for training the fair decision model.

Our 3-Phase Framework (called DeepLF)



- Phase 1: Train a Decision Classifier
  - Train a classifier h_ω as a component of RCGAN.
- Phase 2: Train an RCGAN
  - Train a recurrent conditional GAN to simulate the data generation.
- Phase 3: Train the Long-term Fair Decision Model
  - Train the decision model on the data generated by the recurrent conditional GAN.

The Architecture of RCGAN



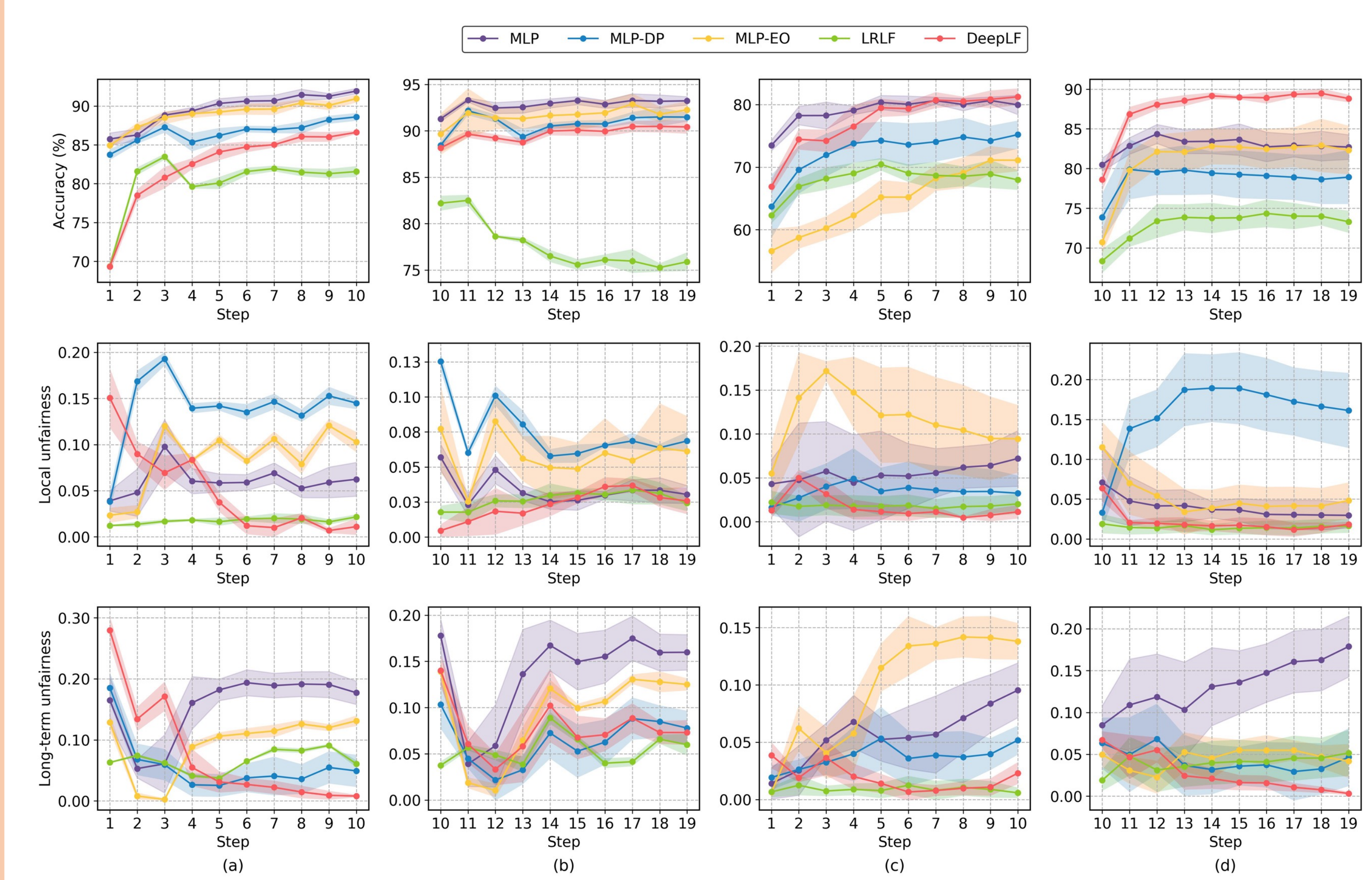
EXPERIMENTS

Baselines

- **MLP**: A MLP without any fairness constraints.
- **MLP-DP**: MLP with DP as fairness constraints.
- **MLP-EO**: MLP with EO as fairness constraints.
- **LRLF**: A logistic regression model with long-term and short-term fairness constraints.

Datasets

- SimLoan (Synthetic)
- Taiwan (Semi-Synthetic)



ACKNOWLEDGEMENT

- This work was supported in part by NSF 1910284, 1946391 and 2142725.