Algorithmic Fairness in Finance: Problems, Methods, and Benchmarks





Zhimeng Jiang[1] Xiaotian Han[1] Chia-Yuan Chang[1] Na Zou[1] Xia Hu[2]

[1] Texas A&M University
[2] Rice university

Tutorial Outline

Part 1: Introduction to Fairness in Finance (Zhimeng and Chia-Yuan)

- Background
- Fairness Definitions
- Methods
  - Pre-/In-/Post-processing overview
  - Showcase of DATA lab research
- Challenges, Insights, and Tools

Part 2 : A Hands-On Example of Fairness in Finance (Xiaotian)

- Fairness Issue in Finance Dataset
- Goal for Financial Fairness: Fairness Metrics
- Hands-on Notebook

Machine Learning are Everywhere in Finance

- Process automation --> Reduced operational cost
- Better productivity --> Increased revenues
- Advanced ML --> Better compliance

Financial Aid

E-commerce

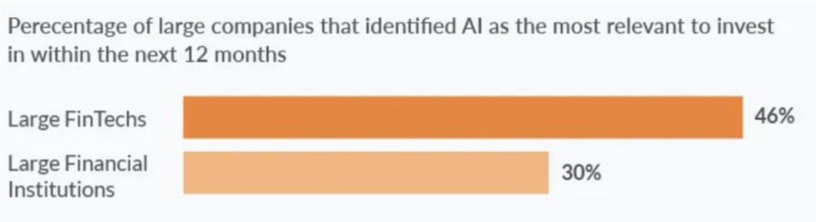


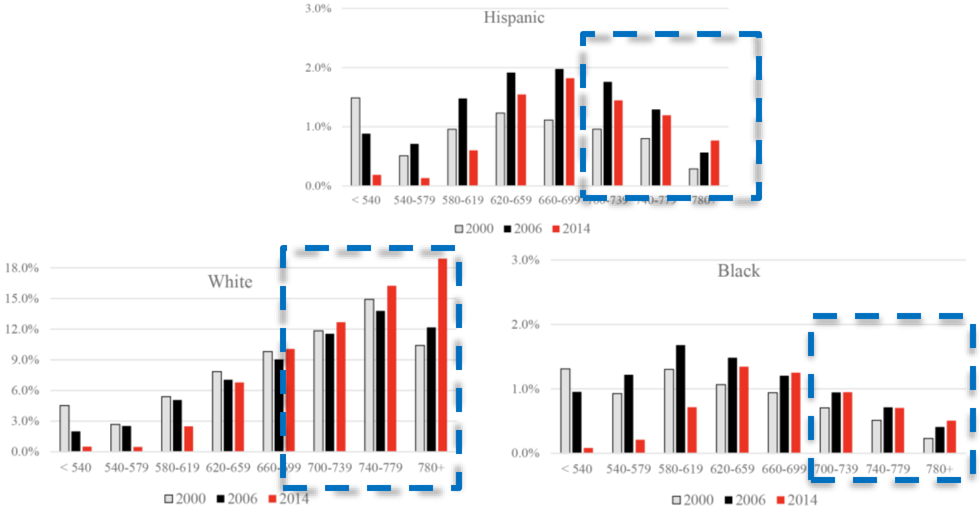
Image source from towards data science:
Machine learning in finance: Why, what & how

Credits Evaluation

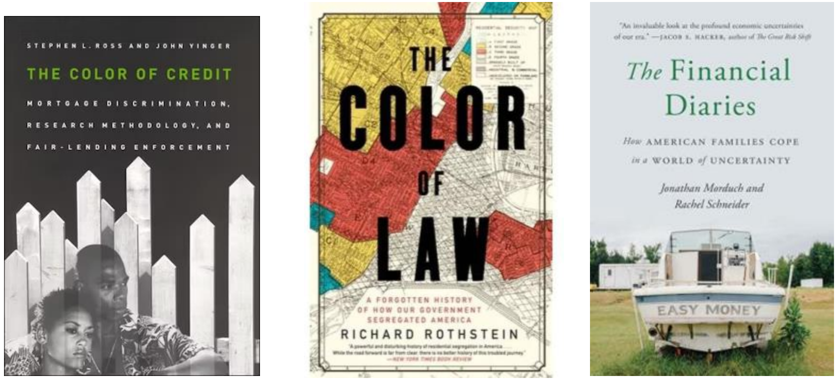
Fairness in Finance

- Foundation laws from the 1960s and 1970s
  - Equal Credit Opportunity Act of 1974
  - Truth in Lending Act of 1968
  - Fair Housing Act of 1968

Credit score distribution varies by race



Bhutta, Neil, and Daniel R. Ringo. *Credit availability and the decline in mortgage lending to minorities after the housing boom*. No. 2016-09-29-2. Board of Governors of the Federal Reserve System (US), 2016.



Some reading on US financial history and sociology

Source from the presentation of Jiahao Chen at NeurIPS 2020

Fairness in Finance

- ML in Finance does need Fairness!


MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV INVESTING CLUB PRO

TECH

A.I. has a discrimination problem. In banking, the consequences can be severe

Ryan Browne
MacKenzie Sigalos

SHARE

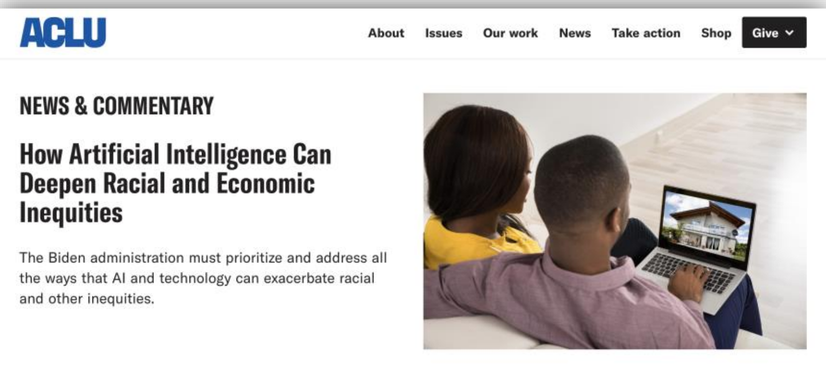
KEY POINTS

- When it comes to banking and financial services, the problem of artificial intelligence amplifying existing human biases can be severe.
- Deloitte notes that AI systems are ultimately only as good as the data they're given: incomplete or unrepresentative datasets could limit AI's objectivity, while biases in development teams that train such systems could perpetuate that cycle of bias.
- Lending is a prime example of where the risk of an AI system being biased against marginalized communities can rear its head, according to former Twitter executive Rumman Chowdhury.

NBC News Brand Studio Mayo Clinic

Groundbreaking advances in healthcare are helping patients at Mayo Clinic who face...

LEARN MORE >

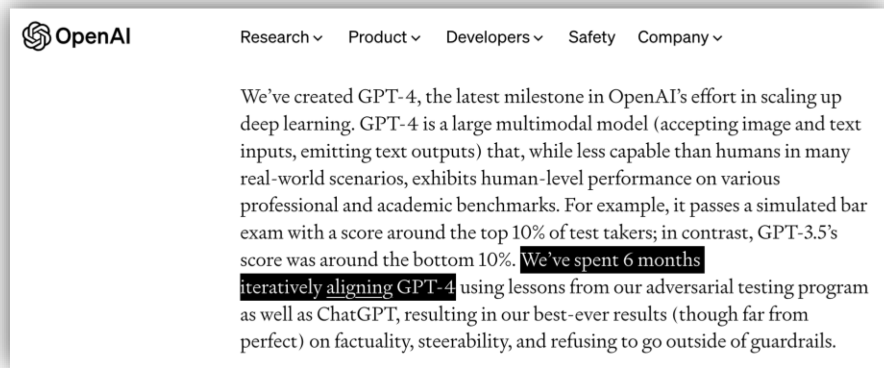

ACLU

About Issues Our work News Take action Shop Give

NEWS & COMMENTARY

How Artificial Intelligence Can Deepen Racial and Economic Inequities

The Biden administration must prioritize and address all the ways that AI and technology can exacerbate racial and other inequities.

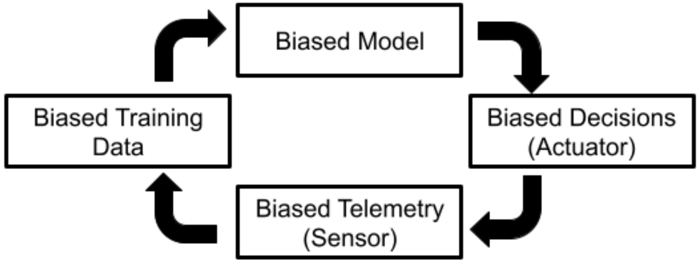

OpenAI

Research Product Developers Safety Company

We've created GPT-4, the latest milestone in OpenAI's effort in scaling up deep learning. GPT-4 is a large multimodal model (accepting image and text inputs, emitting text outputs) that, while less capable than humans in many real-world scenarios, exhibits human-level performance on various professional and academic benchmarks. For example, it passes a simulated bar exam with a score around the top 10% of test takers; in contrast, GPT-3.5's score was around the bottom 10%. We've spent 6 months iteratively aligning GPT-4 using lessons from our adversarial testing program as well as ChatGPT, resulting in our best-ever results (though far from perfect) on factuality, steerability, and refusing to go outside of guardrails.

Bias Life-cycle in Machine Learning

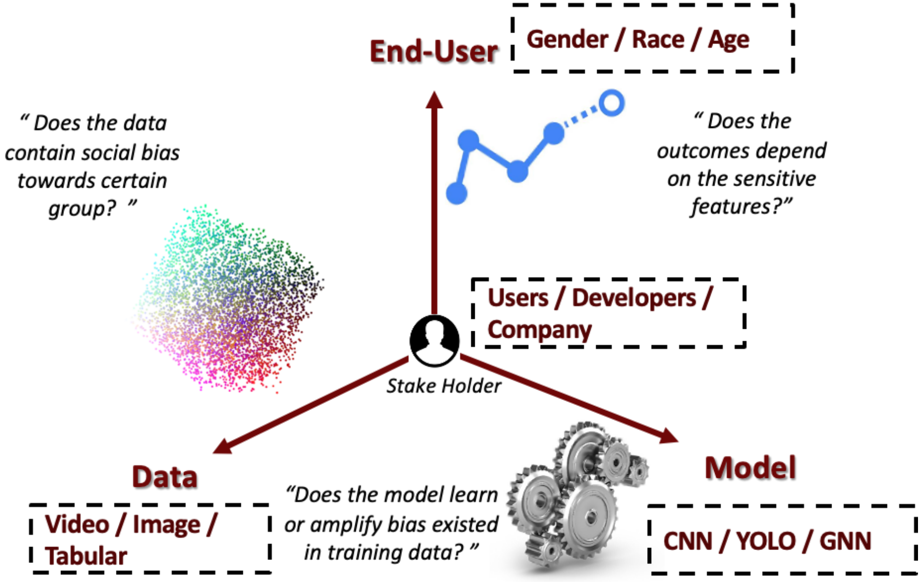
- Inherent bias presented in society
  - Reinforced life-cycle: data – model – prediction
  - A loan example:
    - Elder with higher credit score --> higher approve ratio by model
    - Higher approve ratio by model --> more loan for elder
    - More loan for elder --> higher credit score



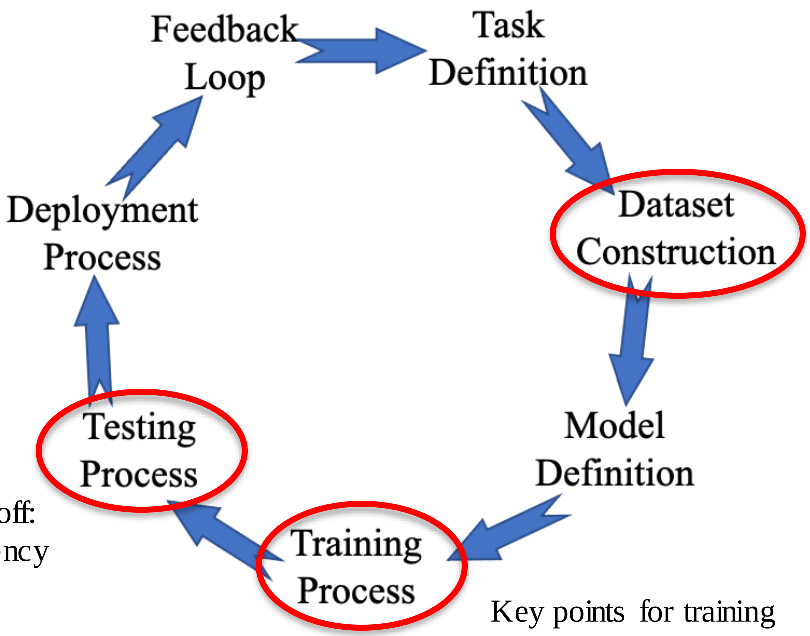
Feedback loop
Image from Medium: link

Fairness in Machine Learning

- Goal: Develop ML/AI systems making decisions with fair treatment
  - Data: human bias leading to biased training data
  - Model: ML model even amplify bias during training
  - End-User: Evaluate outcome bias based on protected attributes



Machine Learning Development Pipeline



Key points for dataset construction
- Bias source: features, label, sample selection
- How to identify data bias?

Key points for evaluation
- Which fairness metric?
- Same test and deployment environment or not?
- Multiple perspectives trade-off: performance, fairness, efficiency

Key points for training
- Bias source: model structure, objective function
- How to mitigate model bias?

Summary

- Fairness is a non-trivial sociotechnical challenge
  - Many types of fairness related to a broad culture context
  - Many fairness definitions
  - Depends on your task definition or collected data
- No free lunch
  - Can't simultaneously satisfy all fairness metrics
  - Fairness v.s. performance
- Bias source
  - Biased training data due to data selection process
  - Biased model due to model structure or training objective
  - Achieving fairness via breaking data – model – prediction life-cycle

Measurements of Fairness

- Group Fairness
  - The difference in model predictions among different sensitive groups

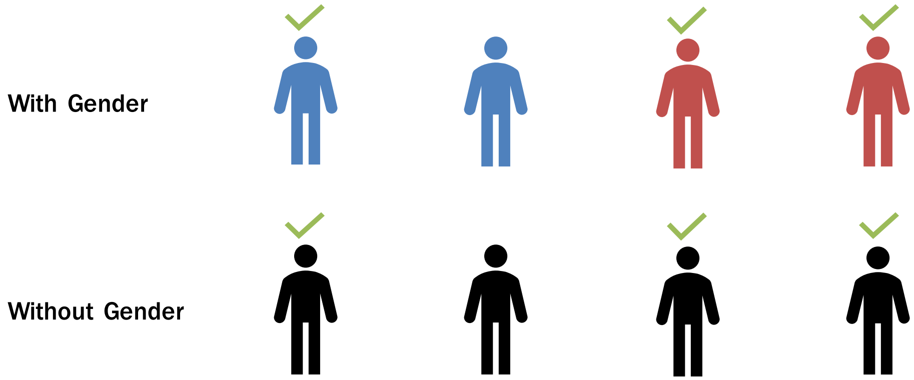
- Individual Fairness
  - The difference in model predictions among similar individuals in different sensitive groups

Measurements of Fairness: **Group Fairness**

- Fairness through Unawareness (FTU)
  - The difference in model predictions between using or not using **sensitive attributes**

$$\mathbb{P}(\hat{y} \mid \mathbf{x}, z) = \mathbb{P}(\hat{y} \mid \mathbf{x})$$

  - Example: Loan Approval Process
    - A loan approval model should make a similar decision with and without sensitive attributes



FTU = 75% - 75%

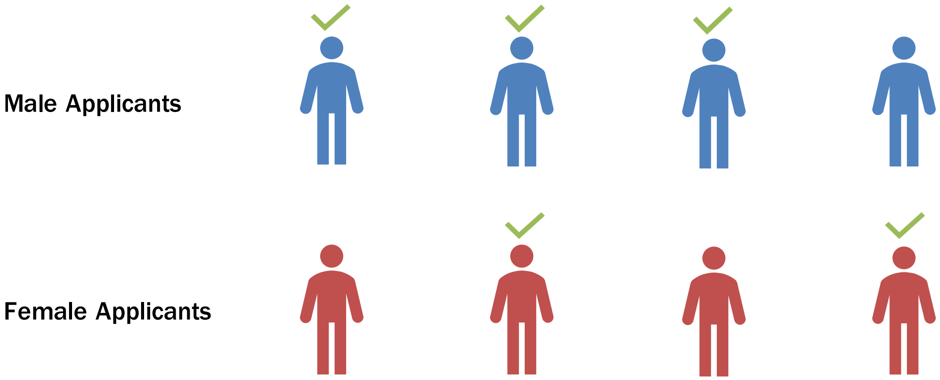
= 0%

Measurements of Fairness: **Group Fairness**

- Demographic Parity (DP)
  - The difference in **positive rates** between different sensitive groups

$$\mathbb{P}(\hat{y} = 1 \,|\, z = a) = \mathbb{P}(\hat{y} = 1 \,|\, z = b)$$

  - Example: Loan Approval Process
    - The difference in the approved applicants from different sensitive groups should be similar



DP = 75% - 50%

= 25%

Measurements of Fairness: **Group Fairness**

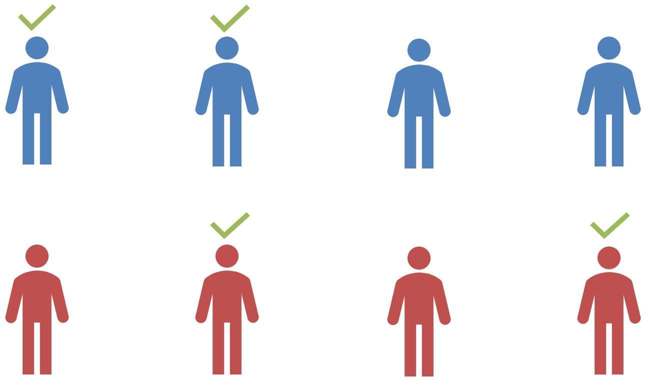
- Equal Opportunity (EO)
  - The difference in **true positive rates** between different sensitive groups

$$\mathbb{P}(\hat{y} = 1 \,|\, y = 1, z = a) = \mathbb{P}(\hat{y} = 1 \,|\, y = 1, z = b)$$

  - Example: Mortgage Lending Process
    - A decision model should approve the similar TPR for eligible majority and minority applicants

Eligible Male Applicants

Eligible Female Applicants

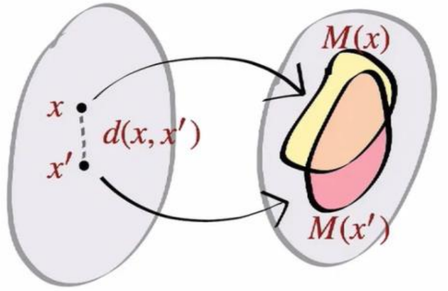


EO = TPR_{Male} - $\text{TPR}_{\text{Female}}$

= 50% - 50%

= 0%

Measurements of Fairness: **Individual Fairness**

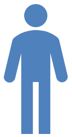


- Fairness through Awareness
  - The difference in model predictions between **similar individuals**

$$D(M(\mathbf{x}), M(\mathbf{x}')) \leq d(\mathbf{x}, \mathbf{x}')$$

  - Example: Credit Scoring Model
    - A credit scoring model should similarly predict two similar clients

Similar Clients



	Financial Behavior: good Income Level: high Credit History: stable Living Area: CA	Financial Behavior: good Income Level: high Credit History: stable Living Area: SF
Credit Level:	**High**	**Hight**

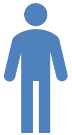
Measurements of Fairness: **Individual Fairness**

- Counterfactual Fairness
  - The difference in model predictions between an individual and its **counterfactual one**

$$\mathbb{P}\left[\hat{y}_{\{\mathrm{z}\leftarrow \mathrm{a}\}} = \mathrm{c} \mid \mathrm{x}, \mathrm{z} = \mathrm{a}\right] = \mathbb{P}\left[\hat{y}_{\{\mathrm{z}\leftarrow b\}} = \mathrm{c} \mid \mathrm{x}, \mathrm{z} = \mathrm{a}\right]$$

  - Example: Credit Scoring Model
    - A credit scoring model should similarly predict a client and its counterfactual one

Real Client



Financial Behavior: good
Income Level: high
Credit History: stable
Living Area: CA

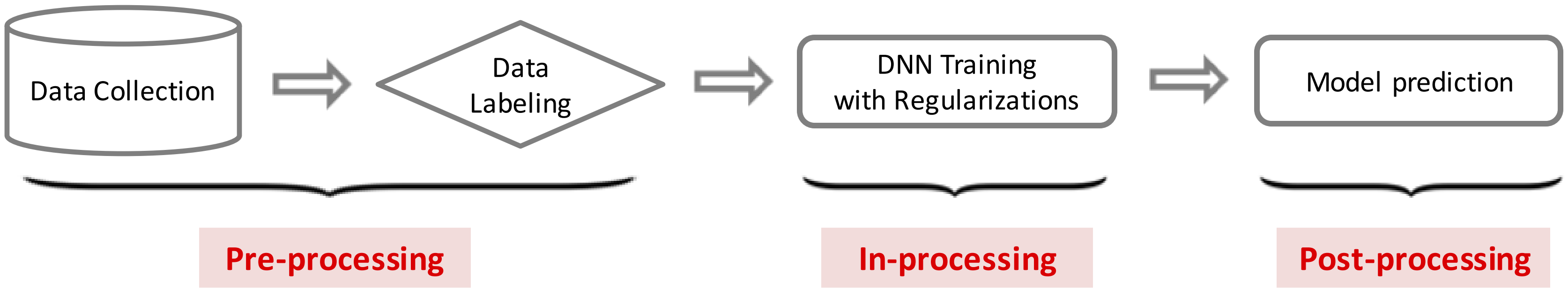
Counterfactual Client

Financial Behavior: good
Income Level: high
Credit History: stable
Living Area: SF

Credit Level: **High** **Hight**

Mitigation Methods

- Three Categories Based on Machine Learning Life-Cycle
  - Pre-processing: debias and increase the quality of training data
  - In-processing: design regularization terms to objective function for learning fair models
  - Post-processing: adjust the outcomes of machine learning models for certain fairness criteria



Mitigation Methods: **Pre-Processing**

- Sampling: upsample minority groups / downsample majority groups
- Data Augmentation: generate synthetic data
  - Example: Co-reference
    - Generate the gender-swapping counterfactual sentences to the training data

Mitigation Methods: **In-Processing**

- Model Constraint
  - Design regularization terms to objective functions based on fairness measurements

$$L(\mathcal{D};\theta) + \lambda||\theta||_2^2 + \underline{\eta R(\mathcal{D};\theta)}$$

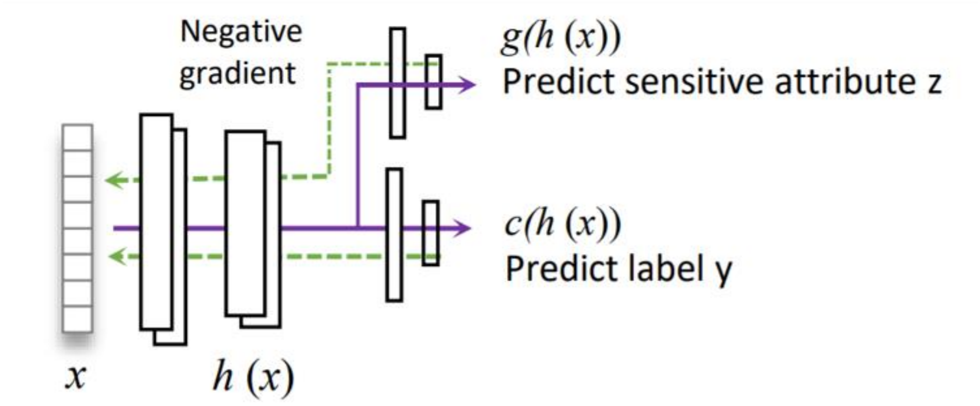
  - Example
    - **Absolute Correlation[1]:** minimize the **absolute correlation** between **Z** and **Y**
    - **Prejudice Index[2]**: minimize the **mutual information** between **Z** and **Y**
    - **Wasserstein fair[3]:** minimize the **Wasserstein distance** between **Z** and **Y**

Z: Sensitive attributes
Y: Model outcomes

[1] Alex Beutel, Jilin Chen, Tulsee Doshi, et al., "Putting Fairness Principles into Practice: Challenges, Metrics, and Improvements." AAAI 2019
[2] Toshihiro Kamishima, Shotaro Akaho, Jun Sakuma, "Fairness-aware Learning through Regularization Approach." IEEE 2011
[3] Ray Jiang, Aldo Pacchiano, Tom Stepleton, Heinrich Jiang, Silvia Chiappa, "Wasserstein Fair Classification." ICML 2020

Mitigation Methods: **In-Processing**

- Adversarial Learning[4]
  - A predictor and an adversarial classifier are learned simultaneously
  - The predictor is trained to accomplish the main task (to predict **Y**)
  - The adversarial classifier is to predict the sensitive attribute **Z**

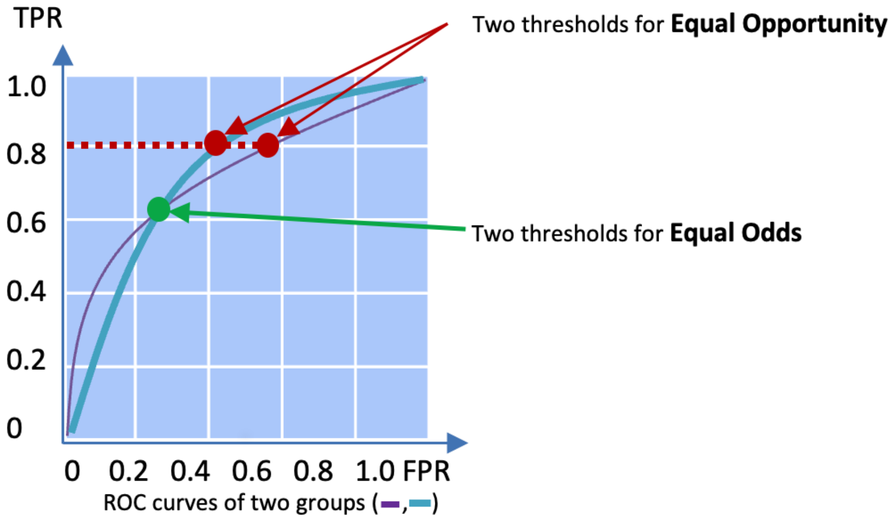


Z: Sensitive attributes
Y: Model outcomes

[4] Brian Hu Zhang, Blake Lemoine, Margaret Mitchell, "Mitigating Unwanted Biases with Adversarial Learning." AAAI 2018

Mitigation Methods: **Post-Processing**

- Different Thresholds for Each Sensitive Group[5]
  - For different fairness measurements, assign a distinctive threshold for each group

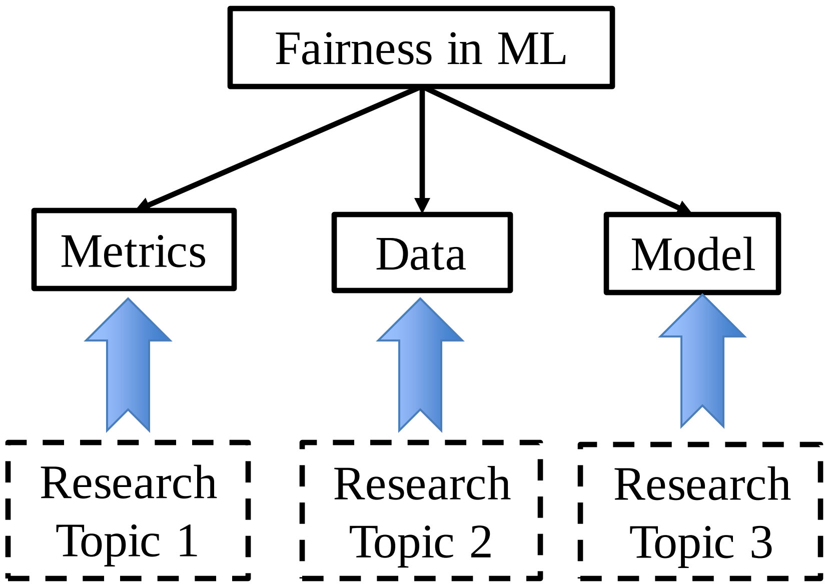


[5] Moritz Hardt, Eric Price, Nathan Srebro, "Equality of Opportunity in Supervised Learning." NeurIPS 2016

Showcases

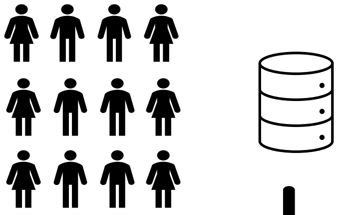
Goal: Develop ML/AI systems that making decisions with fair treatment

- Metrics: Evaluate outcome bias based on protected attributes
- Data: human bias leading to biased training data
- Model: ML model even amplify bias during training



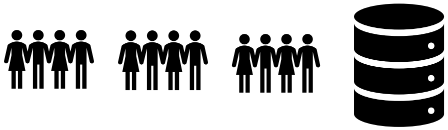
Research Topic 1: Generalized Fairness Metrics

- Existing group fairness metrics are either inapplicable for continuous sensitive attribute or without tractable computation.



Gender	Income
Male	50k
Female	45k

Aggregation

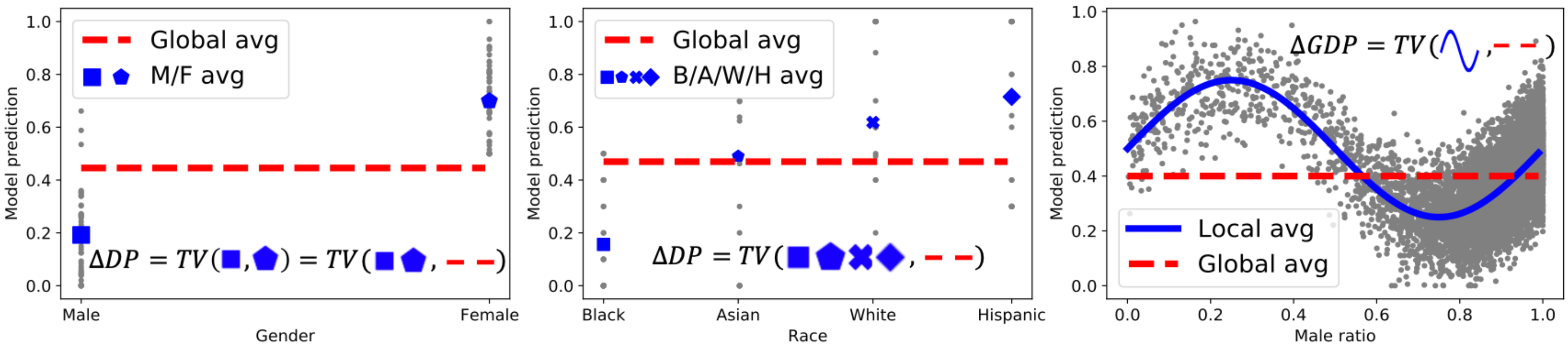


Gender	Income
45%	48.50k
50%	46.2k

Observation: Data aggregation transforms binary sensitive attribute into continuous attributes

GDP Overview

- Demographic parity (DP)[6]: binary sensitive attribute
- Difference w.r.t. DP (DDP)[7]: categorical sensitive attribute
- Generalized DP (GDP): general version for binary/categorical/continuous sensitive attribute
  - local/global difference
  - Local average: average prediction given specific sensitive attribute



[6] Feldman, Michael, et al. "Certifying and removing disparate impact." proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. 2015.
[7] Cho, Jaewoong, et al. "A fair classifier using kernel density estimation." Advances in Neural Information Processing Systems 33 (2020): 15088-15099.

GDP Justifications

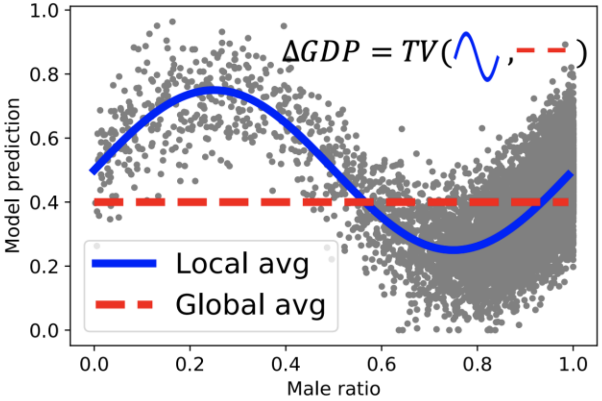
- GDP is a natural extension of DP/DDP for continuous attribute
  - GDP and DP are equivalent except the dataset-dependent coefficient for binary attribute.
  - GDP is weighted DDP for categorical attribute.
- GDP understanding from a probabilistic view
  - Idea case: prediction ⊥ sensitive attribute
    - Joint distribution = Product marginal distribution
  - GDP is a necessary condition for independency
    - GDP ≤ TV distance(joint, product margin)
- GDP regularizer v.s. adversarial debiasing
  - Adversarial debiasing leads to lower GDP

$$\mathcal{L}_{adv}\left(\boxed{g^*}(f(X)), S\right) \geq \Delta GDP.$$

Adversary: Predict sensitive attribute based on NN outputs

GDP Estimations

- Histogram estimation
  - Hard group: consecutive, non-overlapping intervals
  - Internal group average as local average
  - Estimation error v.s #samples: $Err_{hist} = O(N^{-\frac{2}{3}})$
- Kernel estimation
  - Soft group: closer attribute pair, higher weight
  - Normalized weighted average (Nadaraya–Watson kernel estimator)
  - Estimation error v.s #samples: $Err_{kernel} = O(N^{-\frac{4}{5}})$



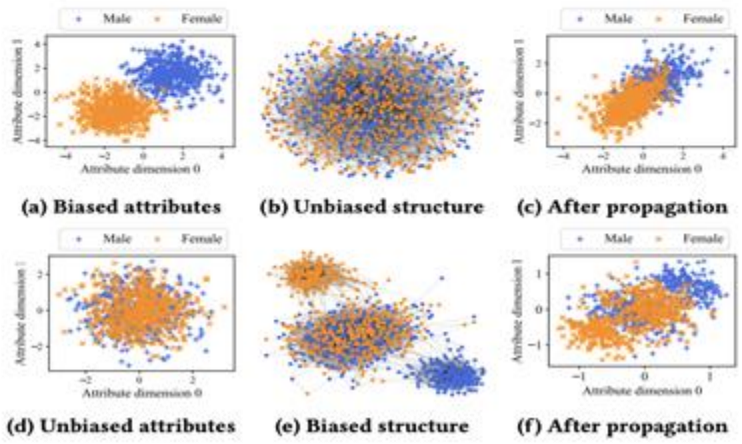
$$\tilde{m}^h(s) = \frac{\sum_{n=1}^{N} \hat{y}_n K(\frac{s_n - s}{h})}{\sum_{n=1}^{N} K(\frac{s_n - s}{h})},$$

$$\tilde{m}^h_{avg} = \frac{\sum_{n=1}^{N} \hat{y}_n}{N}.$$

$$\tilde{\Delta} GDP(h) = \int_0^1 \left| \tilde{m}^h(s) - \tilde{m}^h_{avg} \right| \tilde{p}^h_S(s) \mathrm{d}s.$$

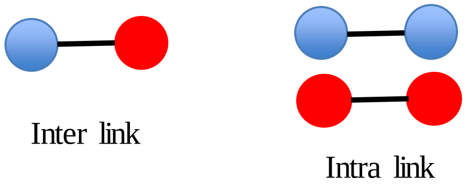
Research Topic 2: Understanding Graph Data Bias

- Understanding the bias in graph neural networks (GNNs)
  - GNNs demonstrate empirical higher prediction bias than peer multilayer perception (MLP)[8] but without theoretical understanding.
  - Bias representation after propagation for bias structure even with unbiased attributes[9].
  - When and Why aggregation enhance the bias?



[8] Dai, Enyan, et al. " Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information." WSDM, 2021.
[9] Dong, Yushun, et al. "Edits: Modeling and mitigating data bias for graph neural networks." WWW, 2022.

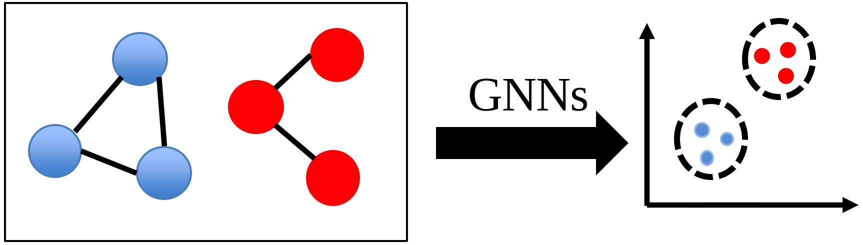
Why Aggregations Suffers?



$$\text{Sensitive Homophily} = \frac{\# \, Intra \; links}{\# \, all \; links}$$

Intuition

- Graph topology with high sensitive homophily coefficient
  - Definition: #sensitive homo links / # links
  - E.g., 95.30% for Pokec-n dataset
  - Higher than label homophily coefficient
- Graph concentration (over-smoothing)
  - More similar representation within demographic group
  - Conditionally happens: no bias for fully over-smoothing



How can we theoretically understand such GNNs behavior?

A Pilot Theoretical Study

Goal: find a sufficient condition of bias enhancement after aggregation

- Synthetic graph data: contexture stochastic block model
  - Topology with intra/inter-connect probability
  - Features with Gaussian Mixture Model
- GCN-like Aggregation
- Bias difference before/after aggregation

When bias enhancement happens

- large sensitive homophily coefficient & node number & connection density
- Balanced demographic size

Topology matters in fair graph learning!

Fair Graph Rewiring

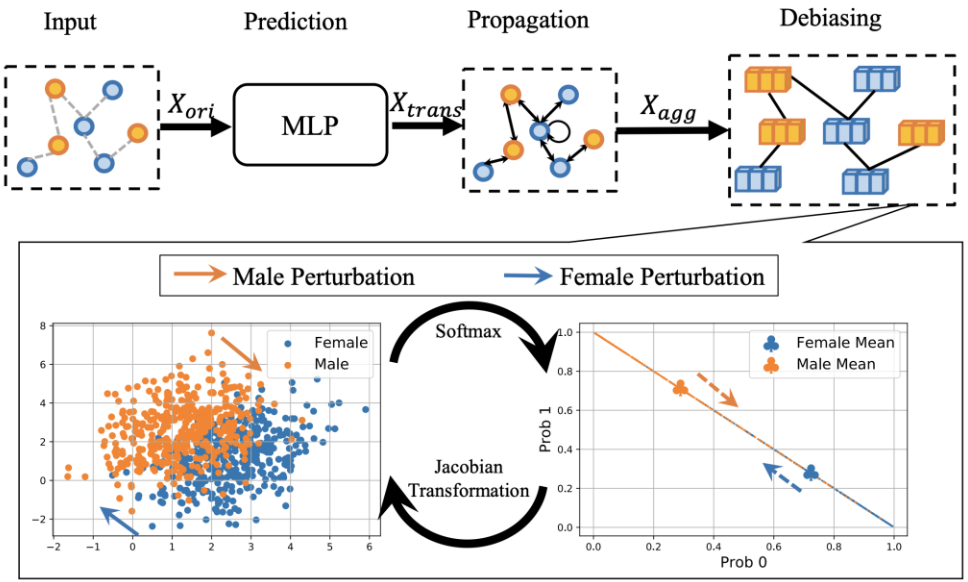
Preprocessing: rewire graph topology to achieve graph fairness

- Large label homophily coefficient
- Low sensitive homophily coefficient
- Low topology perturbation

$$L(\hat{A}|s,y,A) = \underbrace{\boxed{\frac{\|H(ss^T)\odot\hat{A}\|_1}{\|\hat{A}\|_1}}}_{\text{Sensitive Homophily}} - \alpha \underbrace{\boxed{\frac{\|H(yy^T)\odot\hat{A}\|_1}{\|\hat{A}\|_1}}}_{\text{Label Homophily}} + \beta \underbrace{\boxed{\|\hat{A}-A\|_1}}_{\text{Topology Perturbation}}$$

Research Topic 3: Fair Message Passing

- Aggregation operations in GNNs amply bias compared with peer MLP
  - How can we design fair message passing in GNNs?

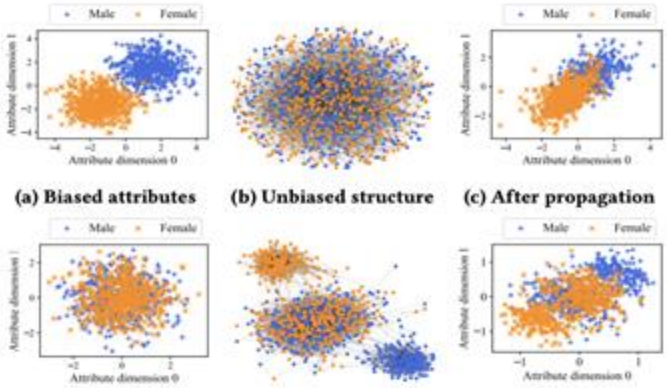


Empirical Observations

- Aggregations in GNNs amplify bias compared with MLP.
  - GNNs > MLP in terms of prediction bias[10]
  - Representation bias after propagation even with unbiased input[11]

Table 2: Results of models w/ and w/o utilizing graph.

Dataset	Metrics	MLP	MLP-e	GCN	GAT
Pokec-z	ACC (%)	65.3 ±0.5	68.6 ±0.3	70.2 ±0.1	70.4 ±0.1
	AUC (%)	71.3 ±0.3	74.8 ±0.3	77.2 ±0.1	76.7 ±0.1
	Δ_{SP} (%)	3.8 ±1.3	6.9 ±1.0	9.9 ±1.1	9.1 ±0.9
	Δ_{EO} (%)	2.2 ±0.7	4.0 ±1.5	9.1 ±0.6	8.4 ±0.6
Pokec-n	ACC (%)	63.1 ±0.4	66.3 ±0.6	70.5 ±0.2	70.3 ±0.1
	AUC (%)	68.2 ±0.3	72.4 ±0.6	75.1 ±0.2	75.1 ±0.2
	Δ_{SP} (%)	3.3 ±0.6	8.7 ±1.0	9.6 ±0.9	9.4 ±0.7
	Δ_{EO} (%)	7.1 ±0.9	9.9 ±0.6	12.8 ±1.3	12.0 ±1.5



(a) Biased attributes (b) Unbiased structure (c) After propagation
(d) Unbiased attributes (e) Biased structure (f) After propagation

[10] Dai, Enyan, et al. " Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information." WSDM, 2021.
[11] Dong, Yushun, et al. "Edits: Modeling and mitigating data bias for graph neural networks." WWW, 2022.

A Unified Optimization Framework

GNNs are graph signal denoising[12]

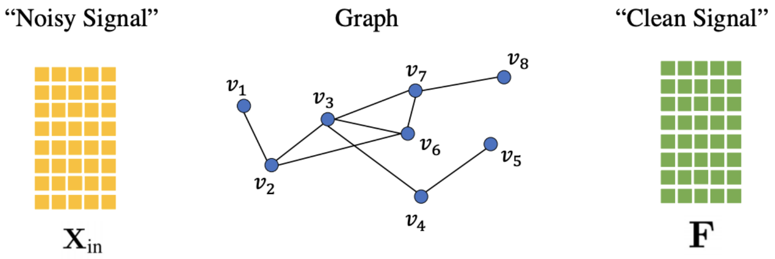
$$\arg\min_{\mathbf{F}} \mathcal{L}(\mathbf{F}) := \|\mathbf{F} - \mathbf{X}_{\text{in}}\|_F^2 + \mathcal{R}(\mathbf{F}, \tilde{\mathbf{L}})$$

Close to the input

Smoothness prior

$$\mathcal{R}(\mathbf{F}, \tilde{\mathbf{L}}) = \lambda \operatorname{tr}(\mathbf{F}^\top \tilde{\mathbf{L}} \mathbf{F}) = \lambda \sum_{(v_i, v_j) \in \mathcal{E}} \left\| \frac{\mathbf{F}_i}{\sqrt{d_i + 1}} - \frac{\mathbf{F}_j}{\sqrt{d_j + 1}} \right\|_2^2$$

Define Prior ⇒ Optimization Solver ⇒ Message Passing



"Nodes are similar to their neighbors"

- GCN $\quad \mathbf{X}_{\text{out}} = \tilde{\mathbf{A}} \mathbf{X}_{\text{in}}$
- PPNP $\quad \mathbf{X}_{\text{out}} = \alpha\big(\mathbf{I} - (1-\alpha)\tilde{\mathbf{A}}\big)^{-1} \mathbf{X}_{\text{in}}$
- APPNP/GCNII $\quad \mathbf{X}^{(k+1)} = (1-\alpha)\tilde{\mathbf{A}}\mathbf{X}^{(k)} + \alpha \mathbf{X}_{\text{in}}$

[12] Ma, Yao, et al. "A unified view on graph neural networks as graph signal denoising." CIKM 2021

Fair Message Passing

Define Prior ⟹ Optimization Solver ⟹ Message Passing

- Objective design

$$\min_{\mathbf{F}} \underbrace{\frac{\lambda_s}{2} tr(\mathbf{F}^T \tilde{\mathbf{L}} \mathbf{F}) + \frac{1}{2} ||\mathbf{F} - \mathbf{X}_{trans}||_F^2}_{h_s(\mathbf{F})} + \underbrace{\lambda_f \boxed{||\mathbf{\Delta}_s SF(\mathbf{F})||_1}}_{h_f\left(\mathbf{\Delta}_s SF(\mathbf{F})\right)} .$$

Fairness prior

- Optimization solver
  - Avoid L1 norm objective via Fenchel conjugate
  - Proximal Alternating Predictor-Corrector Solver[13]

$$\min_{\mathbf{F}} \max_{\mathbf{u}} h_s(\mathbf{F}) + \langle \mathbf{p}, \mathbf{u} \rangle - h_f^*(\mathbf{u})$$

- Fair Message passing

$$\begin{cases} \mathbf{X}_{agg}^{k+1} = \gamma \mathbf{X}_{trans} + (1-\gamma)\tilde{\mathbf{A}}\mathbf{F}^k, & \text{Step ❶} \\ \bar{\mathbf{F}}^{k+1} = \mathbf{X}_{agg}^{k+1} - \gamma \frac{\partial \langle \mathbf{p}, \mathbf{u}^k \rangle}{\partial \mathbf{F}}\Big|_{\mathbf{F}^k}, & \text{Step ❷} \\ \bar{\mathbf{u}}^{k+1} = \mathbf{u}^k + \beta \mathbf{\Delta_s} SF(\bar{\mathbf{F}}^{k+1}), & \text{Step ❸} \\ \mathbf{u}^{k+1} = \min\left(|\bar{\mathbf{u}}^{k+1}|, \lambda_f\right) \cdot sign(\bar{\mathbf{u}}^{k+1}), & \text{Step ❹} \\ \mathbf{F}^{k+1} = \mathbf{X}_{agg}^{k+1} - \gamma \frac{\partial \langle \mathbf{p}, \mathbf{u}^{k+1} \rangle}{\partial \mathbf{F}}\Big|_{\mathbf{F}^k}. & \text{Step ❺} \end{cases}$$

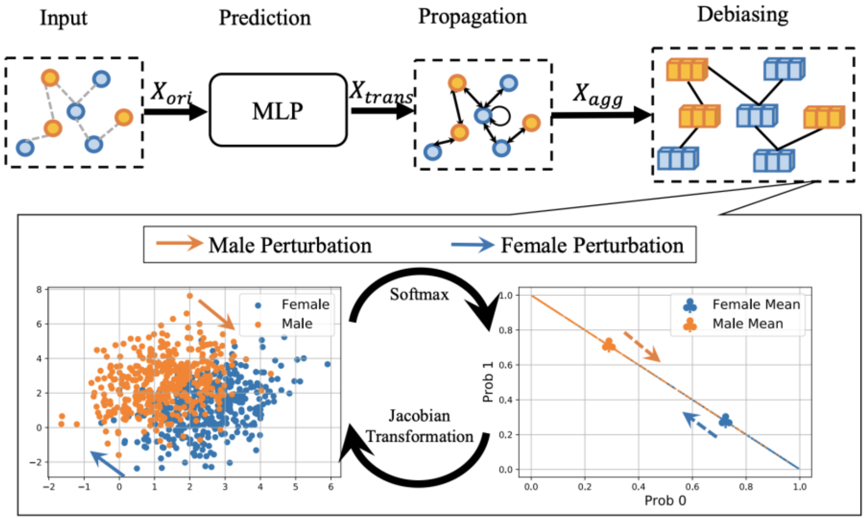
Aggregation with skip connection

Learn and reshape perturbation vector u

[13] Ignace Loris, et al. On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty. Inverse Problems, 27(12):125007, 2011.

Fair Message Passing

- FMP Interpretation
  - Three stages in FMP
  - Four steps in Debiasing
- Efficiency
  - Negligible additional computation
- White-box sensitive attribute usage
  - Explicit usage in FMP
  - Implicit encoding in parameters for fair training

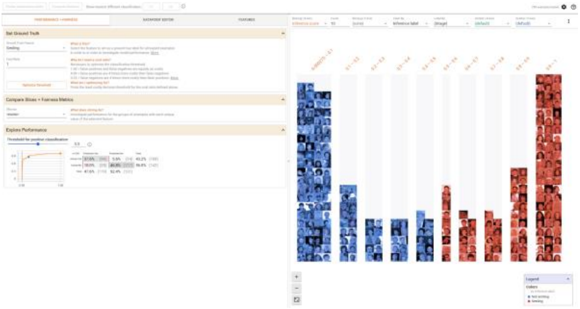


Challenges, Insights, and Tools

- Challenges and Insights
  - Define target fairness for your own task
    - Group fairness, individual fairness or counterfactual fairness?
    - Fairness metric definition
    - Compositional fairness (multiple sensitive attributes)
  - Fairness achievement
    - Data: feature masking, sample selection, data distillation, et al.
    - Model: regularization, adversarial debiasing, reweighting, et al.
    - Prediction: threshold adjustment, calibration
  - Fairness with transparency
    - Bias detection via model interpretation
    - Interpretate fairness algorithms

Challenges, Insights, and Tools

- Tools
  - Google What-if
  - IBM Fairness 360
  - Microsoft Fairlearn
  - DATA Lab FFB



AI Fairness 360

Fairlearn

FFB Fair Fairness Benchmark

A Hands-On Example of Fairness in Finance

- **Fairness Issue in Finance Tasks**
  - Income Prediction
  - Credit Risk Prediction
  - ...
- **A Hands-On Example of Fairness in Finance**
  - Our Proposed Framework: Fair Fairness Benchmark (FFB)
  - A Live Demo

Fairness Issues in Financial Tasks

- Income Prediction
  - Dataset: Adult[1]
  - Sensitive attribute: Gender
- Credit Risk Prediction
- And more…

[1] http://archive.ics.uci.edu/dataset/2/adult

Financial Task: Income Prediction

- Income Prediction
  - Task: Predict whether an individual will earn more or less than $50,000 per year..
  - Dataset: Adult [1]
  - Sensitive attribute: Gender
  - Target: develop a model that accurately predicts the income while ensuring fairness

Prediction

	Age	Workclass	Final Weight	Education	Education Number of Years	Marital-status	Occupation	Relationship	Race	Gender	Capital-gain	Capital-loss	Hours-per-week	Native-country	Income
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K

[1] http://archive.ics.uci.edu/dataset/2/adult

Introducing Fair Fairness Benchmark (FFB)

- The Fair Fairness Benchmark (FFB) is
  - A Pytorch-based framework
  - A set of fair machine learning models
  - Comprehensive fairness evaluation metric

- This benchmark aims to be
  - Minimalistic
  - Hackable
  - Beginner-friendly
  - Reference implementation for researchers

[1] FFB: A Fair Fairness Benchmark for In-Processing Group Fairness Methods, Xiaotian Han, Jianfeng Chi, Yu Chen, Qifan Wang, Han Zhao, Na Zou, Xia Hu
[2] https://github.com/ahxt/fair_fairness_benchmark

A Case Study jupyter on Income Prediction

Q&A





Zhimeng Jiang[1] Xiaotian Han[1] Chia-Yuan Chang[1] Na Zou[1] Xia Hu[2]

[1] Texas A&M University
[2] Rice university