Enhancing Group Fairness In Online Settings Using Oblique Decision Forests

Somnath Basu Roy Chowdhury, Nicholas Monath, Ahmad Beirami, Rahul Kidambi, Avinava Dubey, Amr Ahmed, and Snigdha Chaturvedi
{somnath, snigdha}@cs.unc.edu, {nmonath, beirami, rahulkidambi, avinavadubey, amra}@google.com

UNC NLP ICLR Spotlight

Google Research Google DeepMind

Group Fairness

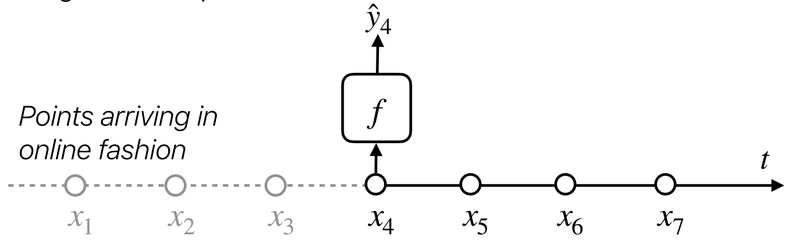
- ML systems often produce unfair decisions against certain groups based on gender, race, religion, etc.
- In this work, we focus on achieving group fairness:

$$\min_f L(f(x), y), \text{ subject to } |\mathbb{E}[f(x|a=0)] - \mathbb{E}[f(x|a=1)]| < \epsilon\,.$$

- Group Fairness ensures different groups receive equal treatment

Group Fairness in Online Settings

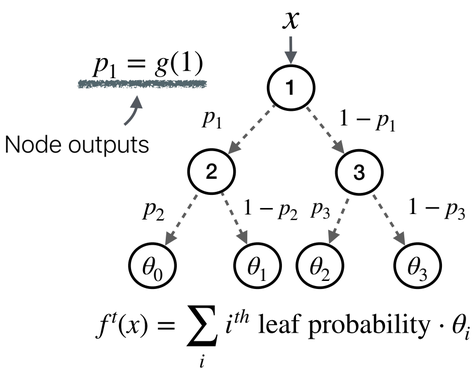
- **Setting**: We focus on the challenging task of achieving fairness in online settings, where points, x_i's, arrive one-at-a-time.
- **Goal**: The system makes a decision for every incoming sample and the overall set of decisions need to be fair
- **Challenge**: Fairness gradient computation requires additional storage and compute.



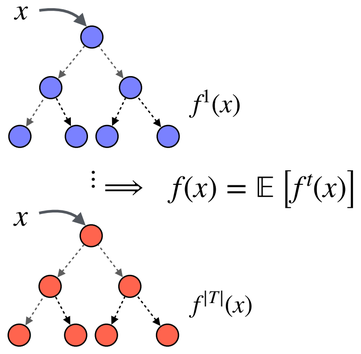
$$\left| \frac{f(x_1|a=0) + \ldots + f(x_n|a=0)}{n} - \mathbb{E}[f(x|a=1)] \right| < \epsilon\,.$$

Challenge: Fairness gradient computation requires storage and multiple passes of f

Aranyani: Fair Oblique Decision Forests



$$f^t(x) = \sum_i i^{th} \text{ leaf probability} \cdot \theta_i$$



- **Prediction function**: Ensemble of oblique decision trees
- **Key Idea**: Apply group fairness constraints on the node decisions, $\mathcal{F}_i$

$$\mathcal{F}_i = |p_i(a=0) - p_i(a=1)|$$

Protected attribute a

- **Implications of node-level constraints**:
  - If $\forall i, \mathcal{F}_i \leq \epsilon$ then overall fairness: $\text{DP} \leq 2^h\epsilon$
  - Easy to estimate $\nabla\mathcal{F}_i$ without storing previous input samples

Theoretical Results

- Estimation error of fairness gradients $\nabla\mathcal{F}_i$ is bounded: $\delta B/2$

δ: Huber constant, B: input bound

- The estimated gradient norm Φ_T (at large T) is bounded by

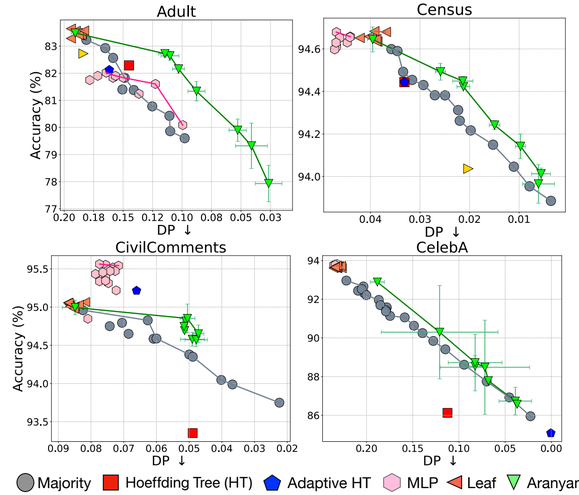
$$\Phi_T \leq \left(\epsilon + 2^{h-2}\lambda^2\delta^2B^2\right)$$

$\epsilon > 0$: small constant

h: tree height, λ: loss hyperparameter

Empirical Results

- We show the effectiveness of Aranyani across *Tabular, Vision, and Language* datasets
- During online learning, at each step we measure the task performance and fairness
- We report average performances at the final step, T



Conclusion

- We propose Aranyani to achieve group fairness in online environments
- Aranyani leverages oblique decision forests for efficient online gradient computation
- Aranyani achieves impressive performance in real-world scenarios

Link to Paper!

brcsomnath/Aranyani