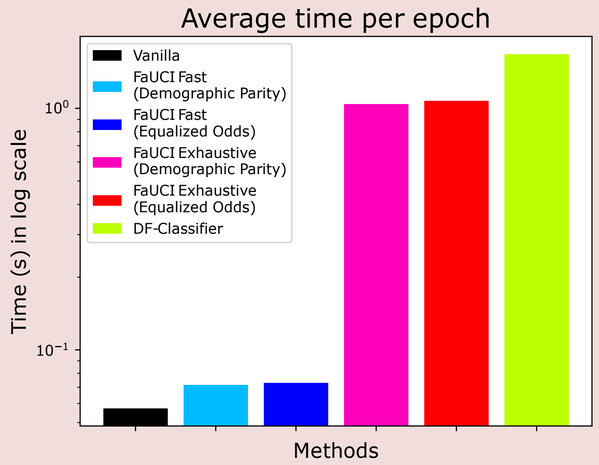
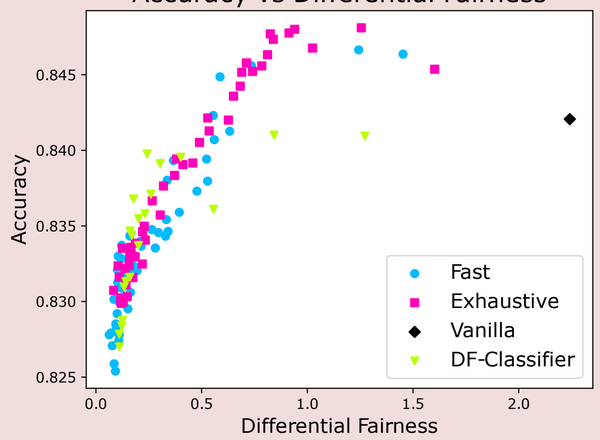
Mitigating Intersectional Fairness: a Practical Approach with FaUCI

Intersectional fairness: account for and mitigate biases experienced across multiple intersecting sensitive features, a.k.a. **subgroups**, of a population (e.g., black women).



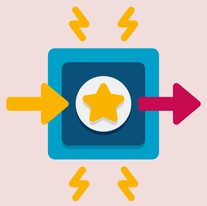


3 major ways to enhance fairness

Pre-processing (data)

In-processing (learning)

Post-processing (corrections)



Solution: instead of computing the fairness function on all the subgroups, just compute it on all the groups of the protected attributes.

Exponential!

$$\mathcal{L}_{h,A}(X,Y) = \mathcal{E}(X,Y) + \lambda * \mathcal{F}_{h,A}(X,Y)$$

Linear!

$$\mathcal{L}_{h,A}(X,Y) = \mathcal{E}(X,Y) + \sum_{i=1}^{n} \lambda_i * \mathcal{F}_{h,A_i}(X,Y)$$

MUCH FASTER!

But does it work?

Issue: regardless of the choice the fairness function, the computational cost of the additional fairness factor is **exponential** in the number of the protected groups. Let A1, ..., An be the protected attributes, each Ai has Gi groups. The theoretical number of possible subgroups is

$$\prod_{i=1}^{n} G_i \rightarrow O(\tilde{G}^n)$$

PROHIBITIVE IN REAL WORLD SCENARIOS!

It is mathematically proven that if we enhance fairnes in the subgroups we also enhance fairness in the groups.

However...

First empirical results show that it is possible to **improve fairness among the subgoups by enhancing the fairness in the groups!**

Matteo Magnini

Alma Mater Studiorum
Università di Bologna

matteo.magnini@unibo.it
http://github.com/matteomagnini

The source code of the experiments with the results is available!

AEQUITAS
unbias AI