Mitigating Intersectional Fairness: a Practical Approach with FaUCI

Matteo Magnini[*] Giovanni Ciatto[*]
Roberta Calegari[*] Andrea Omicini[*]

[*]Dipartimento di Informatica – Scienza e Ingegneria (DISI)
Alma Mater Studiorum – Università di Bologna
{*matteo.magnini*, giovanni.ciatto, roberta.calegari, andrea.omicini}@unibo.it

Workshop on AI bias: Measurements, Mitigation, Explanation Strategies (AIMMES)

20th March, 2024, Amsterdam (NL)

Next in Line...

Context I

Intersectional Fairness:

- looking at multiple intersecting sensitive features
  - gender and ethnicity $\rightarrow$ black women, Hispanic men, etc.
  - age, education, gender $\rightarrow$ old men with high-school diploma, etc.
- and mitigate bias!

Context II

Intersectional Fairness:

- looking at multiple intersecting sensitive features
  - gender and ethnicity → black women, Hispanic men, etc.
  - age, education, gender → old men with high-school diploma, etc.
- and mitigate bias!

Context III

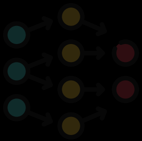
Intersectional Fairness:

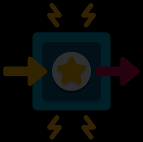
- looking at multiple intersecting sensitive features
  - gender and ethnicity → black women, Hispanic men, etc.
  - age, education, gender → old men with high-school diploma, etc.
- and mitigate bias!

Context IV

3 ways to mitigate bias:

- pre-processing → operations on the dataset
- in-processing → modify the model's (usually a NN) error function
- post-processing → adjust the results of the trained model

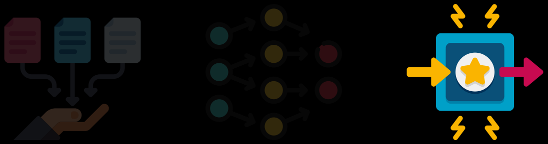




Context V

3 ways to mitigate bias:

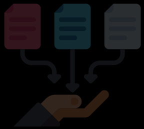
- pre-processing → operations on the dataset
- in-processing → modify the model's (usually a NN) error function
- post-processing → adjust the results of the trained model

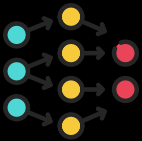


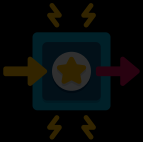
Context VI

3 ways to mitigate bias:

- pre-processing → operations on the dataset
- in-processing → modify the model's (usually a NN) error function
- post-processing → adjust the results of the trained model







Motivation I

The number of subgroups grows exponentially as the number of protected attributes increases!

- a group is an unique value of a predefined attribute (e.g., women)
- the amount of subgroups is the Cartesian product of all possible protected groups
- we need to compute a fairness metric for each of them (e.g., differential fairness)
  - → with many protected attributes the execution time explodes!

Motivation II

The number of subgroups grows exponentially as the number of protected attributes increases!

Next in Line...

Theory

Interesting theory properties:

- reducing bias in subgroups $\implies$ less bias in groups! [Filippi et al., 2023]
- mitigating bias in groups $\not\Longrightarrow$ less bias in subgroups

Idea

Even if by the theory mitigating bias in groups does not necessarily imply less bias in subgroups:

- in actual practice this could happen anyway
- if so, there would be an enormous gain in terms of less computational time cost $\rightarrow$ from exponential to linear!

Next in Line...

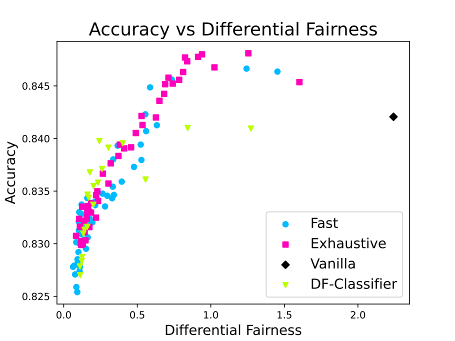
Experiments

We validated our approach on the well known Adult dataset. [Becker and Kohavi, 1996]

- we tried multiple configurations (both linear and exponential in computational cost)
- we also compared to one of the state-of-the-art algorithm, namely DF-Classifier [Foulds et al., 2020]

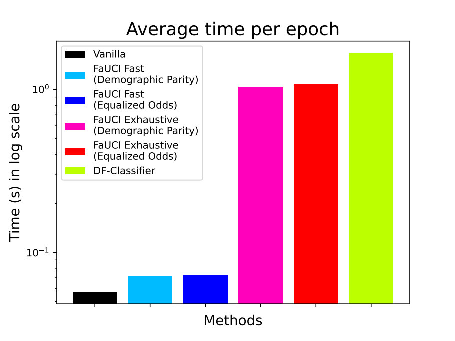
Results & Future works I

Surprisingly the performance of the models trained with methods that have a linear cost are comparable with the other models.



Results & Future works II

As expected, the actual computation time to train for an epoch the NN is way much lower if we consider only the groups.



Results & Future works III

As expected, the actual computation time to train for an epoch the NN is way much lower if we consider only the groups.

Results & Future works IV

Promising direction but:

1. need to test on actually real-world dataset (not Adult)
2. consider more protected attributes
3. validate on other fairness metrics

Mitigating Intersectional Fairness: a Practical Approach with FaUCI

Matteo Magnini[*] Giovanni Ciatto[*]
Roberta Calegari[*] Andrea Omicini[*]

[*]Dipartimento di Informatica – Scienza e Ingegneria (DISI)
Alma Mater Studiorum – Università di Bologna
{*matteo.magnini*, giovanni.ciatto, roberta.calegari, andrea.omicini}@unibo.it

Workshop on AI bias: Measurements, Mitigation, Explanation Strategies (AIMMES)

20th March, 2024, Amsterdam (NL)

References I

[Becker and Kohavi, 1996] Becker, B. and Kohavi, R. (1996).
Adult.
UCI Machine Learning Repository.

[Filippi et al., 2023] Filippi, G., Zannone, S., and Koshiyama, A. S. (2023).
Intersectional fairness: A fractal approach.
CoRR, abs/2302.12683
DOI:10.48550/ARXIV.2302.12683

[Foulds et al., 2020] Foulds, J. R., Islam, R., Keya, K. N., and Pan, S. (2020).
An intersectional definition of fairness.
In *36th IEEE International Conference on Data Engineering, ICDE 2020, Dallas, TX, USA, April 20-24, 2020*, pages 1918–1921. IEEE
DOI:10.1109/ICDE48307.2020.00203