TERRITORI APERTI

Enhancing Fairness in Classification Tasks with Multiple Variables: a Data- and Model-Agnostic Approach

Giordano d'Aloisio, Andrea D'Angelo, Antinisca Di Marco, Giovanni Stilo

University of L'Aquila, Italy

UNIVERSITÀ DEGLI STUDI DELL'AQUILA

COMUNE DELL'AQUILA

CGIL CISL UIL

Outline

- Introduction
- Fairness definitions
- Debiaser for Multiple Variables
- Experimental evaluation
- Conclusion and future works

Introduction

- Bias impacts individuals or groups characterized by a set of legally-protected sensitive attributes (e.g., race, gender, religion, …)
- If not managed, the inequalities reinforced by search and recommendation algorithms can lead to severe *discrimination* and *unfairness*


Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

ON A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden and a friend grabbed the bike and scooter and tried to ride them down the street in the Fort Lauderdale suburb of Coral Springs.

Just as the 18-year-old girls were realizing they were too big for the tiny conveyances — which belonged to a 6-year-old boy — a woman came running after them saying, "That's my kid's stuff." Borden and her friend immediately dropped the bike and scooter and walked away.

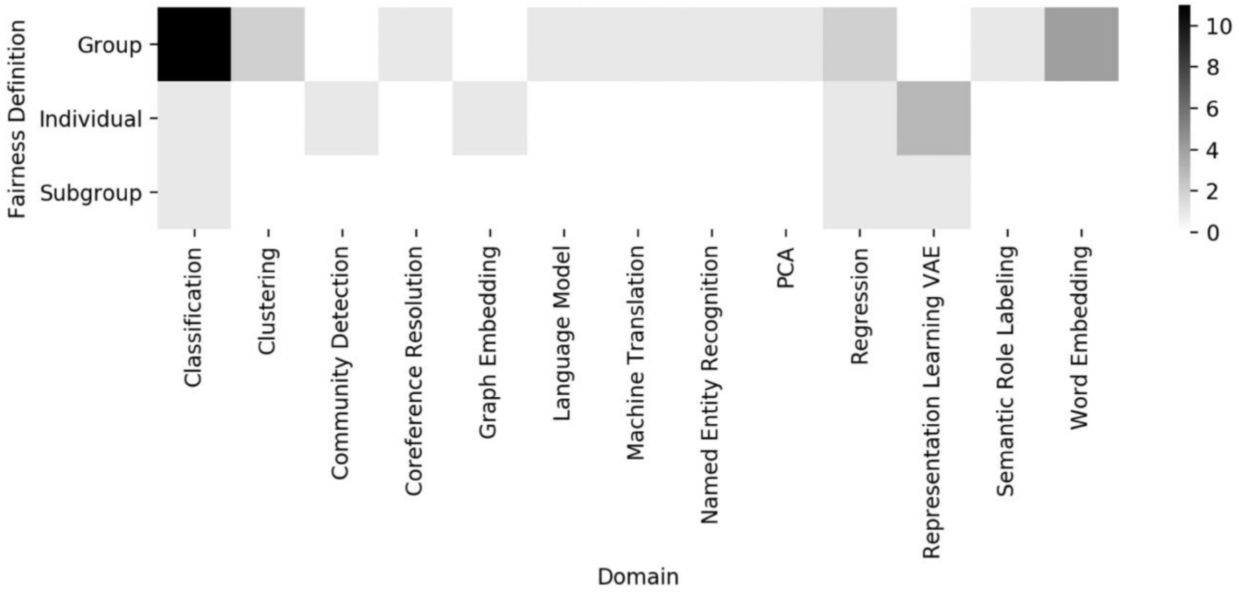
But it was too late — a neighbor who witnessed the heist had already called the police. Borden and her friend were arrested and charged with burglary and petty theft for the items, which were valued at a total of $80.

[Machine Bias — ProPublica]

Motivation

- Over the years many methods have been proposed to mitigate bias in classification domain

- However, we notice that the multi-class classification problem is still not effectively been addressed



Distribution of bias mitigation methods for fairness definition and ML domain, from [1]

For this reason, we present the *Debiaser for Multiple Variables (DEMV)*, a model- and data-agnostic *pre-processing* approach to mitigate bias in binary and multi-class domain with any sensitive variable

[1] Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A. A Survey on Bias and Fairness in Machine Learning. ACM Computational Survey 2021, 54 (6), 1–35.

Fairness Definitions

Statistical (Demographic) Parity (SP)

Independence among the predicted positive label y_p and the sensitive variables $S_1, S_2, \dots, S_n$:

$$P\left(\hat{Y} = y_p \middle| S = 0\right) = P(\hat{Y} = y_p | S = 1)$$

Disparate Impact (DI)

Different formulation of SP which considers the ratio among the two probabilities:

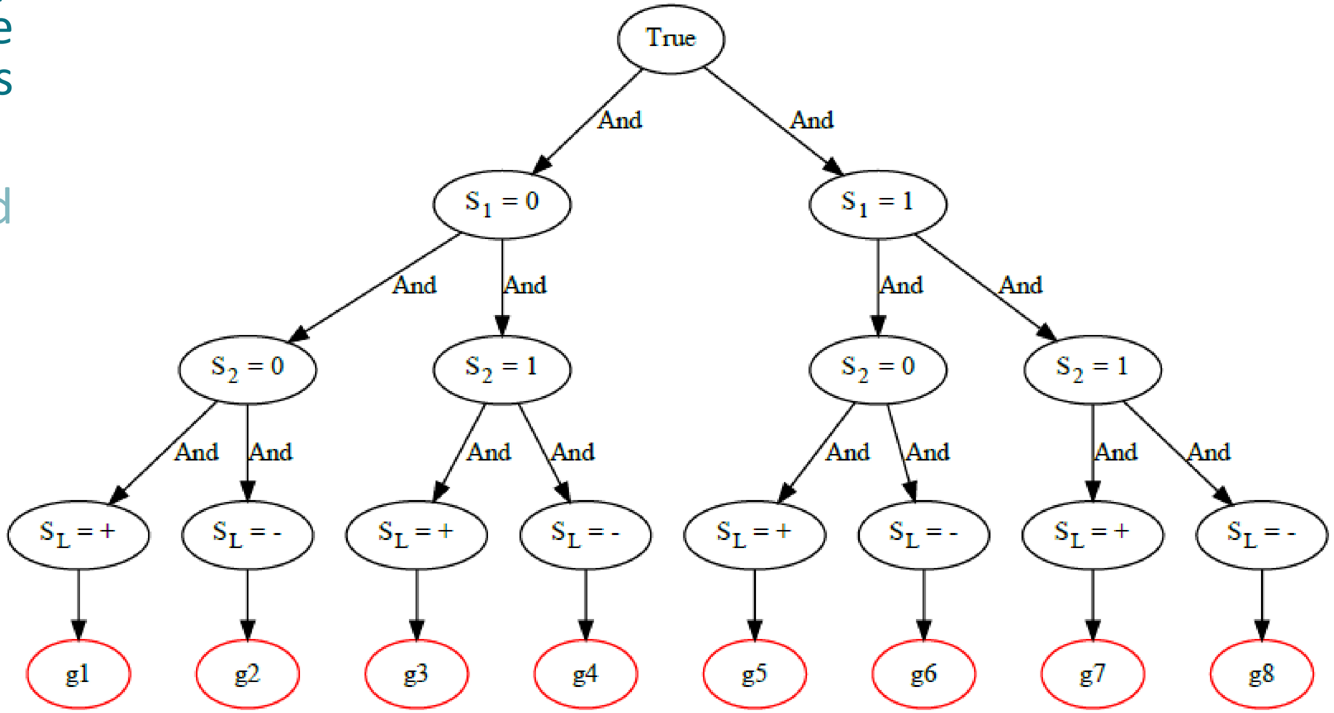
$$\frac{P\left(\hat{Y} = y_p \middle| S = 0\right)}{P\left(\hat{Y} = y_p \middle| S = 1\right)}$$

Debiaser for Multiple Variables (DEMV)

- Identify all the sensitive groups made by all the possible combinations of the sensitive variables' values and label's values
- For each group g, compute its observed (W_{obs}) and expected (W_{exp}) size, then:
  - If ${W_{exp}}/{W_{obs}} > 1$, then:
    - Randomly duplicate an item *i* from *g*
  - Else if ${W_{exp}}/{W_{obs}} < 1$, then:
    - Randomly remove an item *i* from *g*
  - Recompute W_{obs}
  - Repeat until ${W_{exp}}/{W_{obs}} = 1$
- Merge the groups and return the balanced dataset

Debiaser for Multiple Variables (DEMV)

- Identify all the sensitive groups made by all the possible combinations of the sensitive variables' values and label's values
- For each group g, compute its observed (W_{obs}) and expected (W_{exp}) size, then:
  - If $W_{exp}/W_{obs} > 1$, then:
    - Randomly duplicate an item *i* from *g*
  - Else if $W_{exp}/W_{obs} < 1$, then:
    - Randomly remove an item *i* from *g*
  - Recompute W_{obs}
  - Repeat until $W_{exp}/W_{obs} = 1$
- Merge the groups and return the balanced dataset



Debiaser for Multiple Variables (DEMV)

- Identify all the sensitive groups made by all the possible combinations of the sensitive variables' values and label's values
- For each group g, compute its observed (W_{obs}) and expected (W_{exp}) size, then:
  - If ${W_{exp}}/{W_{obs}} > 1$, then:
    - Randomly duplicate an item *i* from *g*
  - Else if ${W_{exp}}/{W_{obs}} < 1$, then:
    - Randomly remove an item *i* from *g*
  - Recompute W_{obs}
  - Repeat until ${W_{exp}}/{W_{obs}} = 1$
- Merge the groups and return the balanced dataset

$$W_{exp} = \frac{|\{X \in D | S = s\}|}{|D|} * \frac{|\{X \in D | L = l\}|}{|D|}$$

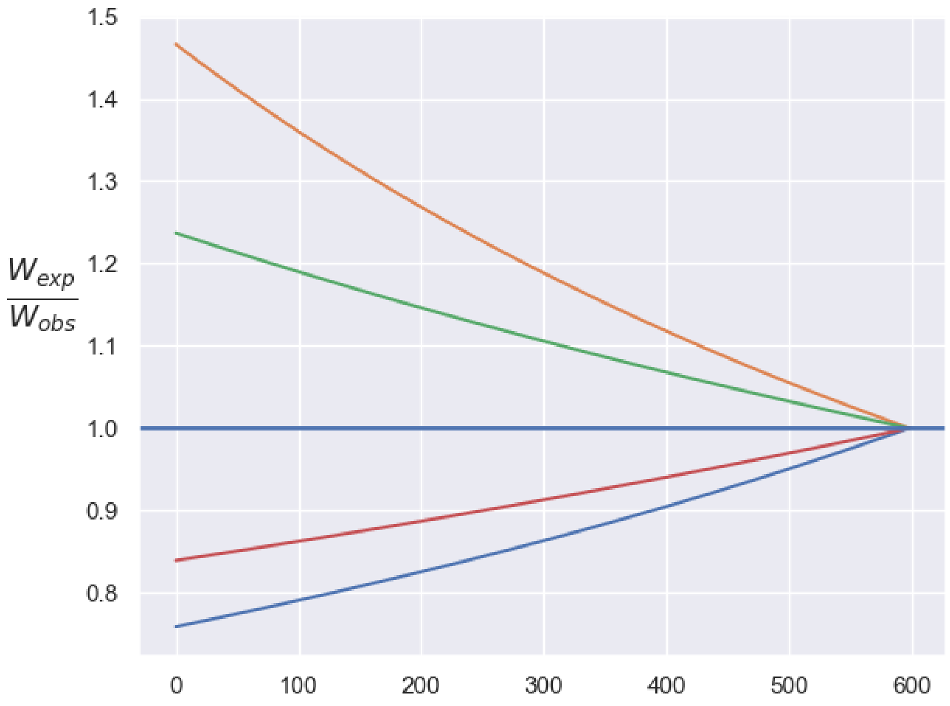
$$W_{obs} = \frac{|\{X \in D | S = s \wedge L = l\}|}{|D|}$$

Where $S = s$ is a generic condition on the sensitive variables' value (binary, discrete or categorical) and $L = l$ is a condition on the label's value

Debiaser for Multiple Variables (DEMV)

- Identify all the sensitive groups made by all the possible combinations of the sensitive variables' values and label's values
- For each group g, compute its observed (W_{obs}) and expected (W_{exp}) size, then:
  - If $W_{exp}/W_{obs} > 1$, then:
    - Randomly duplicate an item *i* from *g*
  - Else if $W_{exp}/W_{obs} < 1$, then:
    - Randomly remove an item *i* from *g*
  - Recompute W_{obs}
  - Repeat until $W_{exp}/W_{obs} = 1$

  sampling
- Merge the groups and return the balanced dataset



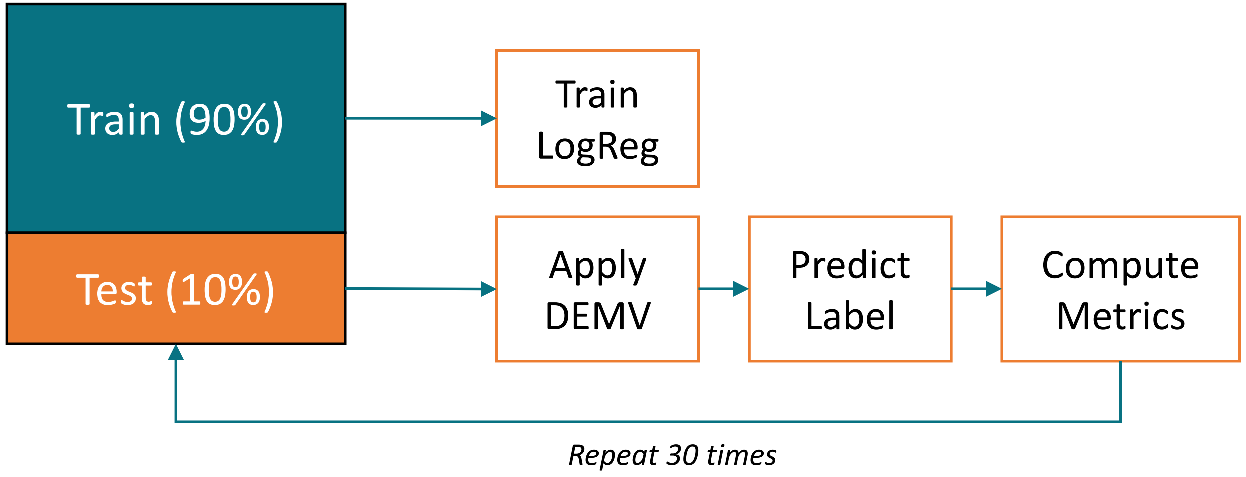
Experimental setting

- We compared our method with the *Exponentiated Gradient* algorithm from [2], using *Statistical Parity* and *Zero One Loss* as constraint for binary and multi-class classifications respectively
- We trained a *Logistic Regression* classifier
- We performed a *10-fold* cross validation and computed the following metrics on the testing set:
  - *Statistical Parity (SP)*
  - *Disparate Impact (DI)*
  - *Zero One Loss (ZO Loss)*
  - *Accuracy (Acc)*

[2] Agarwal, A.; Beygelzimer, A.; Dudik, M.; Langford, J.; Wallach, H. A Reductions Approach to Fair Classification. In *Proceedings of the 35th International Conference on Machine Learning*; PMLR, 2018; pp 60–69.

DEMV evaluation

- Since DEMV has a stochastic behavior in the item's duplication and removal, for train-test fold, we applied DEMV and performed the described metrics 30 times



Employed datasets

- We analyzed DEMV with a heterogeneous set of binary and multi-class datasets from the bias and fairness literature

	Adult	Compas	German	CMC	Crime	Law	Trump	Wine
Scope	Social	Justice	Social	Social	Justice	Education	Social	Food
Instances	30,940	6,167	1,000	1473	1,994	20,427	7,951	6,438
Features	102	399	59	10	100	14	204	13
Type	binary	binary	binary	multi	multi	multi	multi	multi
Sensitive variables	sex race	sex race	sex age	work religion	black hisp	gender race	religion gender	type alcohol
Percentage of sensitive group	5.02%	54.71%	10.50%	64.83%	23.62%	8.42%	30.71%	11.40%

Employed datasets

- We analyzed DEMV with a heterogeneous set of binary and multi-class datasets from the bias and fairness literature

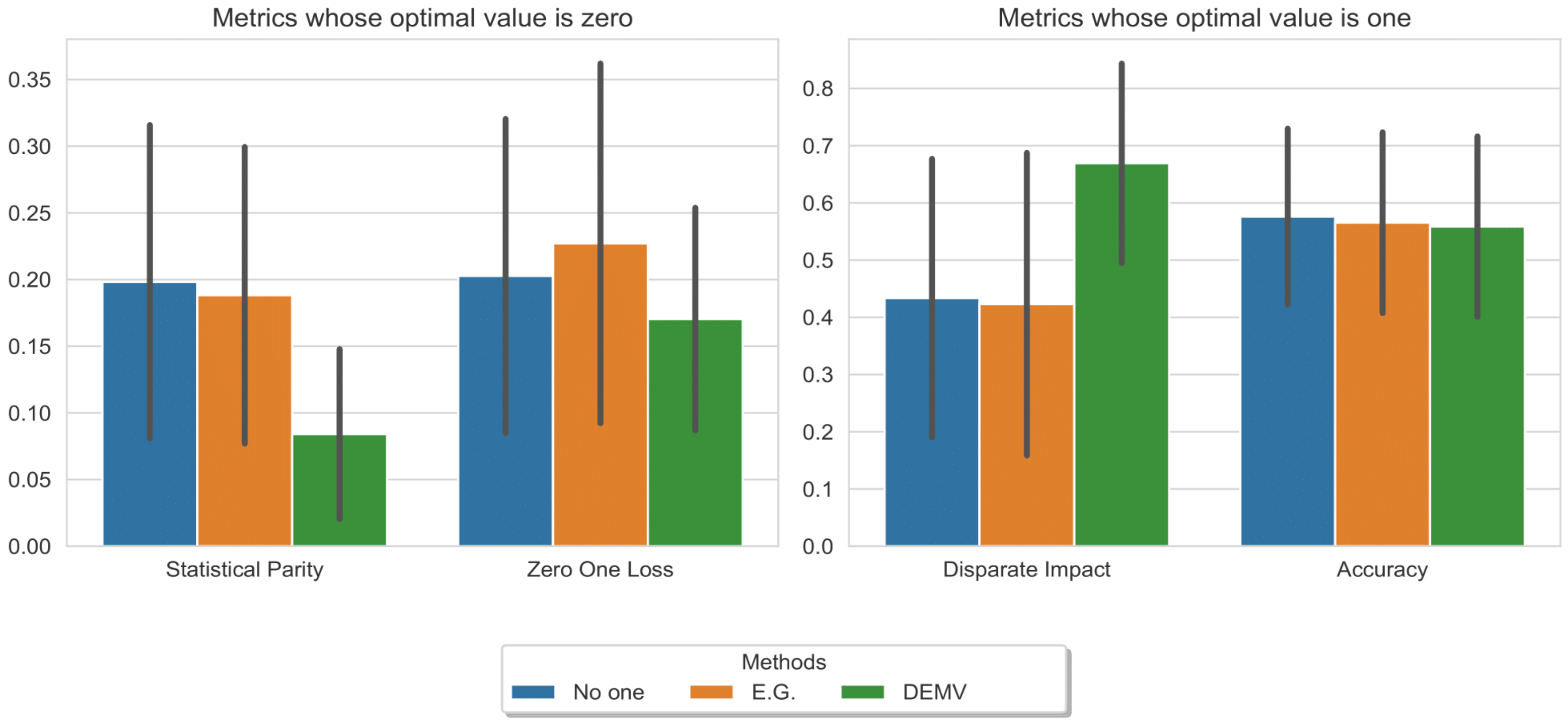
	Adult	Compas	German	CMC	Crime	Law	Trump	Wine
Scope	Social	Justice	Social	Social	Justice	Education	Social	Food
Instances	30,940	6,167	1,000	1473	1,994	20,427	7,951	6,438
Features	102	399	59	10	100	14	204	13
Type	binary	binary	binary	multi	multi	multi	multi	multi
Sensitive variables	sex race	sex race	sex age	work religion	black hisp	gender race	religion gender	type alcohol
Percentage of sensitive group	5.02%	54.71%	10.50%	64.83%	23.62%	8.42%	30.71%	11.40%

Employed datasets

- We analyzed DEMV with a heterogeneous set of binary and multi-class datasets from the bias and fairness literature

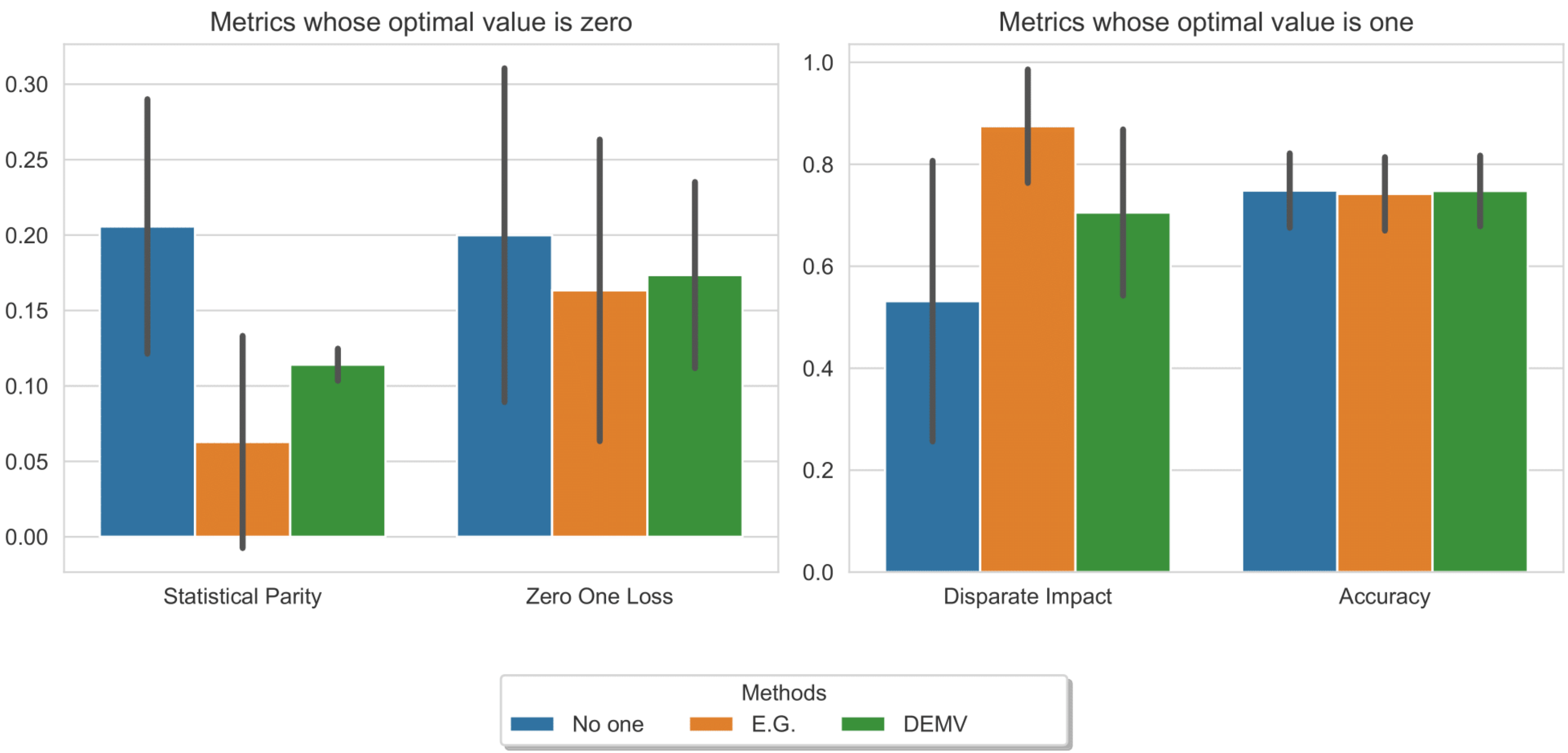
	Adult	Compas	German	CMC	Crime	Law	Trump	Wine
Scope	Social	Justice	Social	Social	Justice	Education	Social	Food
Instances	30,940	6,167	1,000	1473	1,994	20,427	7,951	6,438
Features	102	399	59	10	100	14	204	13
Type	binary	binary	binary	multi	multi	multi	multi	multi
Sensitive variables	sex race	sex race	sex age	work religion	black hisp	gender race	religion gender	type alcohol
Percentage of sensitive group	5.02%	54.71%	10.50%	64.83%	23.62%	8.42%	30.71%	11.40%

Experimental results for multi-class datasets



- Overall mean and standard deviation of the metrics for the biased classifier, EG and DEMV for multi-class datasets

Experimental results for binary datasets



- Overall mean and standard deviation of the metrics for the biased classifier, EG and DEMV for binary datasets

Discussion

Discussion

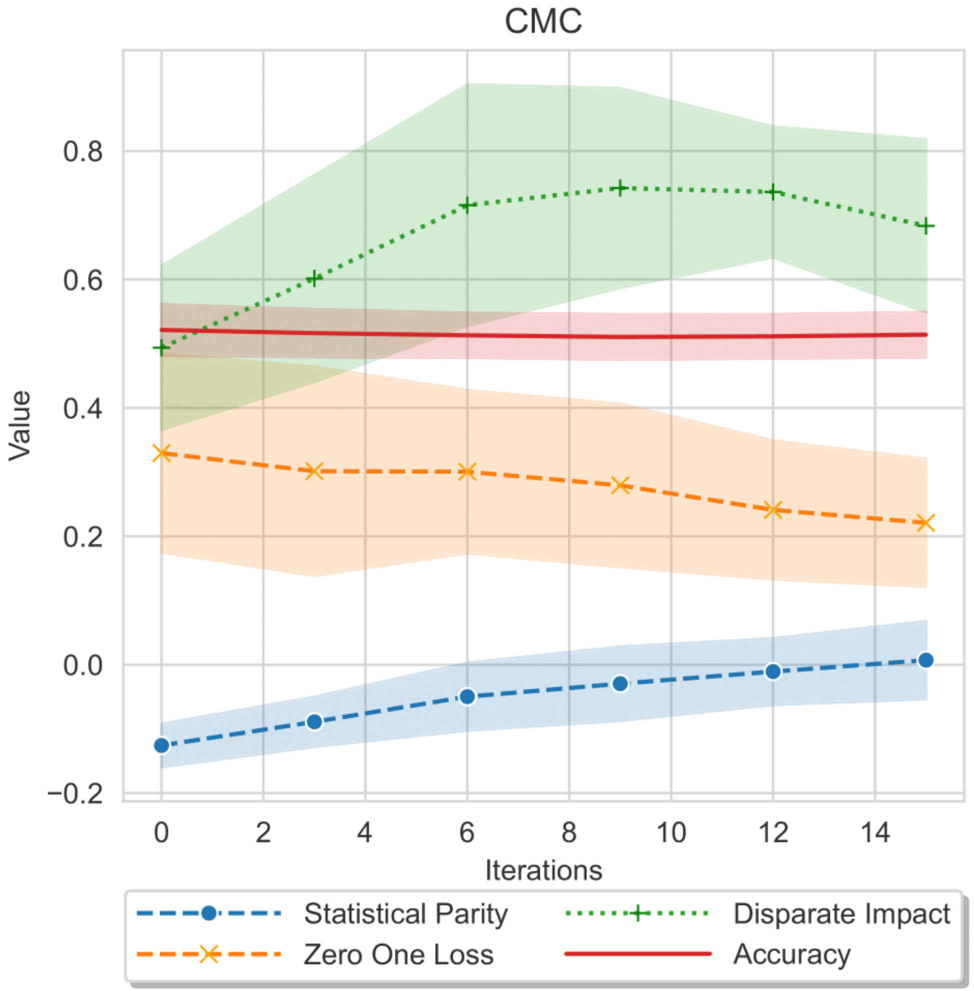
- DEMV is able to improve fairness in multi-class classification domain

Discussion

- DEMV is able to improve fairness in multi-class classification domain
- Concerning binary classification our method has more difficulty improving fairness especially when the starting bias is very high

Discussion

- DEMV is able to improve fairness in multi-class classification domain
- Concerning binary classification our method has more difficulty improving fairness especially when the starting bias is very high
- Finally, we noticed how not always the best fairness is achieved with a complete balancing of the groups



Conclusion and future works

- DEMV is a novel approach, primarily defined for the under explored multi-classification domain
- DEMV is a better strategy to adopt than EG in multi-class tasks
- Performing a complete balancing is not always the optimal solution for all datasets
- DEMV is also able to improve fairness in binary classification. However, as expected, other specifically designed methods may perform better in such cases
- In future, we like to investigate which are the characteristics of the dataset that lead to optimal fairness before a complete balance of the groups
- In addition, we want to widely test DEMV with a different number of sensitive variables, more metrics, more datasets and more baselines

Thank you for your attention!

Source code: https://bit.ly/3E18Q9y

Scan Me