Paper: OmniFair: A Declarative System for Model-Agnostic Group Fairness in Machine Learning. [1]

Jiaxin Liu

Group Reading

November 4, 2021

Overview

1. Introduction
   - Group Fairness Constraints
2. OmniFair
   - Declarative Specification
   - Single Fairness Constraint
   - Multiple Fairness Constraints
3. Experiments

Notations and group fairness constraints

- $x \in \mathbb{R}^d$, $y \in \{0,1\}$ and $D = \{(x_i, y_i)\}_{i=1}^N$
- Goal: $h_\theta(x)$, where $h : \mathbb{R}^d \to \{0,1\}$.

Group Fairness constraints:

- Statistical Parity (SP): $\forall g_i, g_j \in G, \Pr(h(x) = 1 | g_i) \backsimeq \Pr(h(x) = 1 | g_j)$
- False Positive Rate Parity (FPR):
  $\forall g_i, g_j \in G, \Pr(h(x) = 1 | g_i, y = 0) \backsimeq \Pr(h(x) = 1 | g_j, y = 0)$
- False Omission Rate Parity (FOR):
  $\forall g_i, g_j \in G, \Pr(y = 1 | g_i, h(x) = 0) \backsimeq \Pr(y = 1 | g_j, h(x) = 0)$
- Misclassification Rate Parity (MR):
  $\forall g_i, g_j \in G, \Pr(h(x) = y | g_i) \backsimeq \Pr(h(x) = y | g_j)$
- False Negative Rate Parity (FNR), False Discovery Rate Parity (FDR)

Declarative Specification

DEFINITION 1. Fairness Specification (g, f, ε) and Fairness Constraint

A fairness specification is said to be satisfied by a classifier h on D if and only if all induced fairness constraints are satisfied, i.e.,

$$\forall g_i, g_j \in g(D), |f(h, g_i) - f(h, g_j)| \leq \varepsilon$$

where $g(D)$ is the declarative grouping function and $f(h, g)$ is the declarative fairness metric function.

Problem Formulation:

- Given a dataset D, an ML algorithm $\mathscr{A}$
- One or multiple group fairness constraints given by one or multiple (g, f, ε).
- Goal: h_θ that maximizes for accuracy, while satisfying given constraint(s).

Declarative Grouping Function $g(D)$

$g(D)$ is a user-defined function that partitions the inpput data to different groups.

Example 1

A dataset $D = \{t_1, \ldots, t_{10}\}$, where t_4, t_5, t_7 and t_9 are African American and others are Caucasian. The grouping function partitions D as $g(D) = \{AfericanAmerican : [4, 5, 7, 9], Caucasion : [1, 2, 3, 6, 8, 10]\}$.

Fairness Constraint

```
def grouping(Dataset D)
  groups = {}

  # user code here with example
  groups['African-American'] = []
  groups['Caucasian'] = []
  for i in range(D.shape(0)):
     groups[D[i]['race']].append(i)

  return groups
```

Figure: Interface for group function.

Declarative Fairness Metric Function $f(h, g)$

$f(h, g)$ takes a classifier h and a group g as input, and returns $(1 + |g|)$ coefficients that specify how the metric is computed:

$$f(h, g) = \sum_{i \in g} c_i \mathbb{1}(h(x_i) = y_i) + c_0$$

	$c_i \mid y_i = 0$	$c_i \mid y_i = 1$	c_0
MR	$1/\lvert g\rvert$	$1/\lvert g\rvert$	0
SP	$-1/\lvert g\rvert$	$1/\lvert g\rvert$	$\lvert\{i : i{\in}g,\, y_i{=}0\}\rvert/\lvert g\rvert$
FPR	$1/\lvert\{i : i{\in}g,\, y_i = 0\}\rvert$	0	0
FNR	0	$1/\lvert\{i : i{\in}g,\, y_i = 1\}\rvert$	0
FOR	$1/\lvert\{i : i{\in}g, h(x_i){=}0\}\rvert$	0	0
FDR	0	$1/\lvert\{i : i{\in}g, h(x_i) = 1\}\rvert$	0

Figure: Coefficients for different popular group fairness metrics.

Single Fairness Constraint

- Accuracy part: $AP(\theta) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(h_\theta(x_i) = y_i)$
- Fairness part: $FP(\theta) = f(h_\theta, g_1) - f(h_\theta, g_2)$

$$\max_\theta AP(\theta) \tag{1}$$

$$s.t. \quad |FP(\theta)| \leq \varepsilon \tag{2}$$

Lagrangian dual function:

$$\begin{aligned} h(\lambda_1, \lambda_2) &= \max_\theta AP(\theta) + \lambda_1(\varepsilon - FP(\theta)) + \lambda_2(\varepsilon + FP(\theta)) \\ &= \max_\theta AP(\theta) + (\lambda_2 - \lambda_1)FP(\theta) + (\lambda_1 + \lambda_2)\varepsilon \end{aligned}$$

* $h(\lambda_1, \lambda_2)$ provides an upper bound for eq.1 for any $\lambda_1 > 0, \lambda_2 > 0$.

cont.

$$\max_{\theta} AP(\theta) + (\lambda_2 - \lambda_1)FP(\theta) + (\lambda_1 + \lambda_2)\varepsilon \tag{3}$$

$$\max_{\theta} AP(\theta) + \lambda FP(\theta) \tag{4}$$

LEMMA 1.

Assume the primal equation is feasible and let θ^* be an optimal solution to primal problem, then

1. for any $\lambda_1, \lambda_2 > 0, h(\lambda_1, \lambda_2) \geq AP(\theta^*)$; and under strong duality assumption, $\min_{\lambda_1>0,\lambda_2>0} h(\lambda_1, \lambda_2) = AP(\theta^*)$
2. for any $\lambda_1, \lambda_2 > 0$, let $\tilde{\theta}$ be an optimal solution to eq.(3), then there exists $\lambda \in \mathbb{R} \quad (i.e. \lambda = \lambda_2 - \lambda_1)$ such that $\tilde{\theta}$ also optimizes eq.(4); and under stong duality assumption, there exists λ such that θ^* optimizes eq.(4).

$$\max_{\theta} AP(\theta) + \lambda FP(\theta) \tag{5}$$

Steps:

1. enumerate all possible λ values;
2. find optimal θ for every λ
3. pick the one that has the maximal $AP(\theta)$.

How to solve eq. (5) for a given λ? (step 2)

$$\begin{aligned}
&\max_{\theta} AP(\theta) + \lambda FP(\theta) \\
&= \max_{\theta} \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_i + \lambda \left(\sum_{i \in g_1} (c_i^{g_1} \mathbb{1}_i + c_0^{g_1}) - \sum_{i \in g_2} (c_i^{g_2} \mathbb{1}_i + c_0^{g_2}) \right) \\
&= \max_{\theta} \frac{1}{N} \sum_{i=1}^{N} w_i(\lambda, h_\theta) \mathbb{1}(h_\theta(x_i) = y_i)
\end{aligned}$$

How to get weight $w_i(\lambda)/w_i(\lambda, h_\theta)$?

weight	metric	$w_i \mid y_i = 0, g_1$	$w_i \mid y_i = 1, g_1$	$w_i \mid y_i = 0, g_2$	$w_i \mid y_i = 1, g_2$
$w_i(\lambda)$	MR	$1 + \lambda N/\lvert g_1\rvert$	$1 + \lambda N/\lvert g_1\rvert$	$1 - \lambda N/\lvert g_2\rvert$	$1 - \lambda N/\lvert g_2\rvert$
	SP	$1 - \lambda N/\lvert g_1\rvert$	$1 + \lambda N/\lvert g_1\rvert$	$1 + \lambda N/\lvert g_2\rvert$	$1 - \lambda N/\lvert g_2\rvert$
	FPR	$1 - \lambda N/\lvert\{i : i \in g_1, y_i = 0\}\rvert$	1	$1 + \lambda N/\lvert\{i : i \in g_2, y_i = 0\}\rvert$	1
	FNR	1	$1 - \lambda N/\lvert\{i : i \in g_1, y_i = 1\}\rvert$	1	$1 + \lambda N/\lvert\{i : i \in g_2, y_i = 1\}\rvert$
$w_i(\lambda, h_\theta)$	FOR	$1 - \lambda N/\lvert\{i : i \in g_1, h(x_i) = 0 \rvert$	1	$1 + \lambda N/\lvert\{i : i \in g_2, h(x_i) = 0\}\rvert$	1
	FDR	1	$1 - \lambda N/\lvert\{i : i \in g_1, h(x_i) = 1\}\rvert$	1	$1 + \lambda N/\lvert\{i : i \in g_2, h(x_i) = 1\}\rvert$

Figure: Weights for different popular group fairness metrics.

- For $w_i(\lambda)$: any black-box ML algorithm.
- For $w_i(\lambda, h_\theta)$:
  - if $\lambda_2 - \lambda_1 \leq \delta = 0.00001$, then $h_{\theta_1|\lambda_1}(x_i) = h_{\theta_2|\lambda_2}(x_i)$.
  - So $w_i(\lambda_2, h_{\theta_2}) \backsimeq w_i(\lambda_2, h_{\theta_1})$.
  - Starting from $\lambda = 0$ and taking small incremental steps δ to estimate weights.

How to tune the hyperparameter λ?

LEMMA 2. Monotonicity for Single Fairness Constraint

Consider two values λ_1 and λ_2, where $\lambda_1 < \lambda_2$, and let θ_1 and θ_2 denote two optimal solutions given λ_1 and λ_2. Then, the following proprieties hold, where $AP(\theta_1), AP(\theta_2), FP(\theta_1), FP(\theta_2)$ are evaluated on the same input training set D:

$$FP(\theta_1) \leq FP(\theta_2)$$
$$AP(\theta_1) \geq AP(\theta_2) \quad where \lambda_2 \geq \lambda_1 \geq 0$$
$$AP(\theta_1) \leq AP(\theta_2) \quad where 0 \geq \lambda_2 \geq \lambda_1$$

* ML model has the maximum accuracy when $\lambda = 0$.

Algorithm for tuning λ

Algorithm 1 Tuning Single λ

Input: Dataset D, a fairness constraint (g, f, ε), an ML Algorithm $\mathcal{A}$

Output: A fair ML model h_θ

```
θ_0 ← apply 𝒜 with w_i(0)
flag ← false if (weights are parameterized by θ) else true
if |FP(θ_0)| ≤ ε then return h_{θ_0}
if FP(θ_0) > 0 then
    change the order of g_1 and g_2 in FP
λ_l ← 0 and λ_u ← 1
if flag = true then
    λ_l, λ_u ← ExponentialSearch(λ_l, λ_u)
else
    λ_l, λ_u ← LinearSearch(λ_l, δ) // e.g. δ = 0.001
while λ_u − λ_l ≥ τ do // τ → 0; e.g. τ = 0.0001
    λ_m ← (λ_u + λ_u)/2
    if flag = true then
        θ_m ← apply 𝒜 with w_i(λ_m)
    else
        θ_m ← apply 𝒜 with w_i(λ_m, h_{θ_l})
    if FP(θ_m) < −ε then λ_l ← λ_m
    else λ_u ← λ_m
return h_{θ_m}

function ExponentialSearch(λ_l, λ_u)
    θ_u ← apply 𝒜 with w_i(λ_u)
    while FP(θ_u) < −ε do
        λ_l ← λ_u
        λ_u ← 2 × λ_u
        θ_u ← apply 𝒜 with w_i(λ_u)
    return λ_l, λ_u

function LinearSearch(λ_l, δ)
    λ_u ← λ_l + δ
    θ_u ← apply 𝒜 given λ_u
    while FP(θ_u) < −ε do
        λ_l ← λ_u
        θ_l ← θ_u
        λ_u ← λ_l + δ
        θ_u ← apply 𝒜 with w_i(λ_u, θ_l)
    return λ_l, λ_u
```

Multiple Fairness Constraints

$$\max_{\theta} AP(\theta)$$
$$s.t. \quad |FP_i(\theta)| \leq \varepsilon \qquad \forall i \in \{1, \ldots, k\}$$

Lagrangian dual function is:

$$h(\Lambda_1, \Lambda_2) = \max_{\theta} AP(\theta) + \sum_{i=1}^{k} \lambda_{1,i}(\varepsilon - FP_i(\theta)) + \sum_{i=1}^{k} \lambda_{2,i}(\varepsilon + FP_i(\theta))$$
$$= \max_{\theta} AP(\theta) + \sum_{i=1}^{k} (\lambda_{2,i} - \lambda_{1,i}) FP_i(\theta) + (\lambda_{1,i} + \lambda_{2,i})\varepsilon$$

where $\Lambda_1 =< \lambda_{11}, \lambda_{12}, \ldots, \lambda_{1k} >$ and $\Lambda_2 =< \lambda_{21}, \lambda_{22}, \ldots, \lambda_{2k} >$.

cont.

$$\max_\theta AP(\theta) + \sum_{i=1}^{k} \lambda_i FP_i(\theta)$$

$$= \max_\theta \frac{1}{N} \sum_{i=1}^{N} w_i(\Lambda, h_\theta) \mathbb{1}(h_\theta(x_i) = y_i)$$

LEMMA 4. Marginal Monotonicity for Multiple Fairness Constraints

Consider two settings Λ_1 and Λ_2, that differ only in the j^{th} dimension, namely, $\Lambda_1[j] < \Lambda_2[j]$ and $\Lambda_1[i] < \Lambda_2[i]$ for all $i \neq j$. Let θ_1 and θ_2 denote the optimal solution given Λ_1 and Λ_2. Then

$$FP_j(\theta_1) \leq FP_j(\theta_2)$$

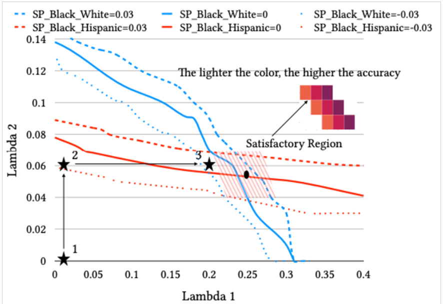
where $FP_j(\theta_1)$ and $FP_j(\theta_2)$ are evaluated on the same training set D.

Algorithm for tuning Λ

Algorithm 2 Hill-Climbing

Input: Dataset D, a set of k fairness constraint, and an ML Algorithm $\mathcal{A}$
Output: A fair ML model h_θ

```
Λ ← [0, ··· , 0]
θ ← apply 𝒜 with w_i(0, −)
while ∃ i, s.t. |FP_i(θ)| > ε_i do
    i ← argmax_k ||FP_k(θ)| − ε_k|
    θ ← call Algorithm 1 to tune for the i-th fairness constraint, while
    fixing Λ[j], ∀j ≠ i to their current values
return Λ if (Λ ∈ intersection of all satisfactory regions) else "Not
found after 5k iterations"
```



Dataset and Experimental settings

- Dataset:

Dataset	# Rows	# Attrs	Sens. Attr	Task
Adult [18]	48842	18	sex	To predict if Income > 50k
Compas [4]	11001	10	race	To predict recidivism
LSAC [45]	27477	12	race	To predict if bar exam is passed
Bank [32]	30488	20	age	To predict if marketing works

  Split each dataset to 60% training, 20% validation, and 20% test.

- ML algorithms:
  - Logistic Regression (LR)
  - Random Forest (RF)
  - XGBoost (XGB)
  - Neural Networks (NN)

Experiments

	COMPAS					Adult					LSAC					Bank				
Algorithm	LR	RF	XGB	NN	CMA-ES	LR	RF	XGB	NN	CMA-ES	LR	RF	XGB	NN	CMA-ES	LR	RF	XGB	NN	CMA-ES
OmniFair	-1.2%	-0.8%	-0.7%	-1.2%	NA(2)*	-2.1%	-1.9%	-1.7%	-1.7%	NA(2)*	-0.3%	-0.3%	-0.4%	-0.1%	NA(2)*	-0.1%	-0.3%	-0.2%	-0.1%	NA(2)*
Kamiran et al. [28]	-2.5%	-1.3%	-1.2%	-1.5%	NA(2)*	-2.7%	-2.3%	-1.8%	-1.9%	NA(2)*	-0.4%	-5.6%	-2.2%	-0.4%	NA(2)*	+0.1%	-1.2%	-0.2%	-0.3%	NA(2)*
Calmon et al. [11]	-1.8%	-0.5%	-0.3%	-0.9%	NA(2)*	-3.7%	-3.1%	-3.0%	-2.4%	NA(2)	NA(1)	NA(1)	NA(1)	NA(1)	NA(2)*	NA(1)	NA(1)	NA(1)	NA(1)	NA(2)*
Zafar et al. [47]	-0.9%	NA(2)	NA(2)	NA(2)	NA(2)	-2.2%	NA(2)	NA(2)	NA(2)	NA(2)	-0.2%	NA(2)	NA(2)	NA(2)	NA(2)	-0.1%	NA(2)	NA(2)	NA(2)	NA(2)
Celis et al. [12]	NA(1)	NA(2)	NA(2)	NA(2)	NA(2)	NA(1)	NA(2)	NA(2)	NA(2)	NA(2)	NA(1)	NA(2)	NA(2)	NA(2)	NA(2)	NA(1)	NA(2)	NA(2)	NA(2)	NA(2)
Agarwal et at. [3]	-2.4%	-1.2%	-2.0%	-1.8%	NA(2)*	-2.8%	-2.2%	-2.0%	-2.0%	NA(2)*	-0.6%	-5.8%	-0.2%	-0.5%	NA(2)*	-0.1%	-0.3%	-0.0%	-0.1%	NA(2)*
Thomas et at. [43]	NA(2)	NA(2)	NA(2)	NA(2)	-1.1%	NA(2)	NA(2)	NA(2)	NA(2)	-1.7%	NA(2)	NA(2)	NA(2)	NA(2)	-0.4%	NA(2)	NA(2)	NA(2)	NA(2)	-0.1%

Figure: Accuracy drop compared with no fairness constraints when $\varepsilon = 0.03$ under SP.

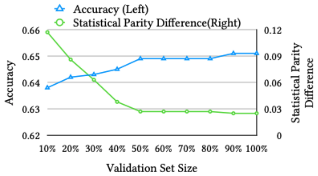


Figure: Ablation study for the validation size on the COMPAS set.

Experiments

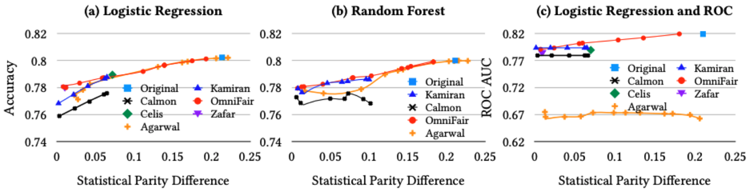


Figure: The trade-off between fairness metric and accuracy on Adult dataset.

ε	Accuracy	SP	FNR
Baseline	63.9%	0.233	0.180
0.01	N/A	N/A	N/A
0.02	N/A	N/A	N/A
0.03	63.6%	0.03	0.044
0.04	63.4%	0.016	0.035
0.05	63.3%	0.028	0.007
0.06	62.7%	0.057	0.032

Figure: Enforcing SP and FNR on COMPAS dataset.

[1] Hantian Zhang et al. "OmniFair: A Declarative System for Model-Agnostic Group Fairness in Machine Learning". In: *Proceedings of the 2021 International Conference on Management of Data*. 2021, pp. 2076–2088.