Fairness with Adaptive Weights

Junyi Chai
chai28@purdue.edu

Xiaoqian Wang
joywang@purdue.edu

July 12, 2022

Purdue University
Elmore Family School of Electrical and Computer Engineering

Introduction

As automated decision making systems are widely applied in social fields, fairness has become an arising concern in machine learning society.

PURDUE UNIVERSITY
Elmore Family School of Electrical and Computer Engineering

Fair classification

Much of literature on fairness focuses on specified fairness metrics.

However, relaxations of fairness metrics could be too relaxed to achieve expected improvement.

Our goal:group balance to mitigate representation bias and error-prone reweighing

PURDUE UNIVERSITY
Elmore Family School of Electrical and Computer Engineering

Method

Equal reweighing:

$$\min_{\theta} \sum_{i=1}^{N} \frac{c}{N_a} L_\theta(y_i, \hat{y}_i).$$

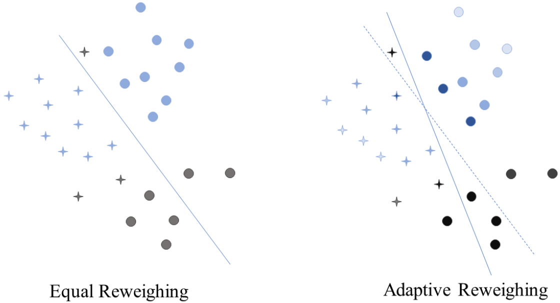

Figure: Demonstration of our method.

Problem formulation

Error-prone reweighing:

$$\min_{\theta} \max_{w} \sum_{i=1}^{n'} w_i L_\theta(y_i, \hat{y}_i) \quad s.t. \quad w^T 1 = c, w \geq 0.$$

Problem formulation

$$\max_{w} \sum_{i=1}^{n'} w_i L(y_i, \hat{y}_i) - \alpha \|w\|_2^2 \ \ s.t. \ \ w^T 1 = c, w \geq 0 \,. \qquad (1)$$

PURDUE UNIVERSITY
Elmore Family School of Electrical and Computer Engineering

Theoretical analysis

Closed-form solution of 1:

$$w_i^* = \max(\frac{l_i - \lambda}{2\alpha}, 0), \quad i = 1, 2, ..., d',$$

where λ is the Lagrange multiplier.

PURDUE UNIVERSITY

Elmore Family School of Electrical and Computer Engineering

Theoretical analysis

Theorem

Consider a classifier f_θ with parameter θ such that $\hat{y}_i = f_\theta(x_i)$. Given the adaptive weight w^* by optimizing Problem (1), under the L_1-norm loss or the cross-entropy loss for $L(y_i, \hat{y}_i)$, the following fairness metrics

- *Disparate mistreatment:*

$$\sum_s \Big(|p(\hat{y} \neq y | y = 1, s) - p(\hat{y} \neq y | y = 1)| + |p(\hat{y} \neq y | y = 0, s) - p(\hat{y} \neq y | y = 0)| \Big)$$

- *Equal opportunity:*

$$\sum_s \left(|p(\hat{y} \neq y | y = 1, s) - p(\hat{y} \neq y | y = 1)| \right)$$

are upper bounded by our weighted loss up to a multiplicative constant.

PURDUE UNIVERSITY
Elmore Family School of Electrical and Computer Engineering

Experiments

Table 3. Experimental results on COMPAS dataset.

Method	Baseline	Reweighing	Undersampling	Oversampling	ASR	Postprocessing	Covariance	Ours
Accuracy	65.23±1.39	62.24±2.47	63.34±2.41	63.50±2.42	63.75±1.27	63.42±1.14	**64.11±1.46**	63.41±1.35
Disparate Impact	22.29±4.76	9.13±3.16	8.45±2.68	8.55±2.83	2.31±0.25	2.33±0.10	7.36±1.03	**1.82±0.11**
Disparate TPR	21.14±7.14	6.46±2.14	9.32±3.86	7.02±3.44	1.07±0.33	1.06±0.16	3.38±0.71	**1.02±0.09**
Disparate TNR	17.41±3.72	19.11±3.22	5.77±1.73	5.25±1.40	1.14±0.21	1.20±0.21	10.28±2.33	**0.24±0.17**

Table 6. Experimental results of nonlinear classifier on COMPAS dataset.

Method	Baseline	Reweighing	Undersampling	Oversampling	ASR	Postprocessing	Covariance	Ours
Accuracy	64.17±1.13	61.18±1.78	62.76±2.26	62.35±2.13	63.17±1.21	63.14±1.16	**63.64±1.31**	63.23±1.64
Disparate Impact	21.37±5.24	10.17±2.27	8.83±2.69	8.67±3.12	2.41±0.31	3.24±0.11	7.43±1.22	**2.23±0.87**
Disparate TPR	22.21±8.17	6.85±2.13	8.86±3.11	7.44±2.57	1.82±0.46	1.31±0.17	3.13±0.76	**1.16±0.08**
Disparate TNR	17.64±3.46	18.85±4.41	5.41±1.68	6.13±1.25	1.71±0.43	1.24±0.23	11.47±2.63	**0.69±0.34**

Purdue University
Elmore Family School of Electrical and Computer Engineering

Experiments

Table: Experimental results on Law school dataset.

Method	MSE	SP
Baseline	0.114±0.003	15.20±4.34%
Oversampling	0.152±0.004	9.62±3.17%
Undersampling	0.163±0.002	8.57±4.52%
FWB	0.141±0.004	**2.13**±**0.13**%
Our method	**0.135**±**0.004**	2.16±0.19%

Table: Experimental results on CRIME dataset.

Method	MSE	SP
Baseline	0.037±0.003	50.63±6.75%
Oversampling	0.052±0.004	21.37±7.73%
Undersampling	0.047±0.006	19.42±6.63%
FWB	**0.042**±**0.004**	12.10±1.19%
Our method	0.043±0.004	**11.47**±**1.63**%

PURDUE UNIVERSITY
Elmore Family School of Electrical and Computer Engineering

Experiments

Fairness-accuracy trade-off:

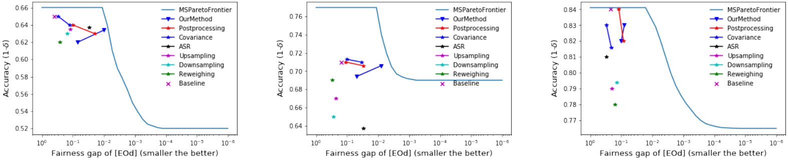

Figure: Pareto frontier on COMPAS, German credit and Adult datasets.

PURDUE UNIVERSITY
Elmore Family School of Electrical and Computer Engineering

Experiments

Robustness:

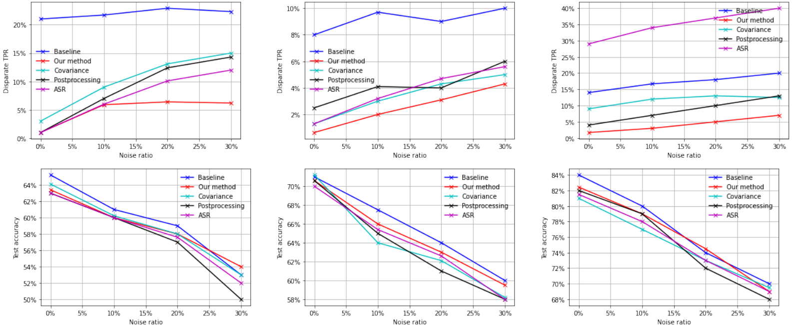

Figure: Change of accuracy and fairness under different noise ratio on COMPAS, German credit and Adult datasets.

Summary

Balance between different groups

Sample-level reweighing method

Close-form solution for weight assignment

Theoretical property in terms of convergence

Fairness guarantee

Thank you

Elmore Family School of Electrical and Computer Engineering