FIFA: Making Fairness More Generalizable in Classifiers Trained on Imbalanced Data

Kyungseon Lee, Choeun Kim, Hankyo Jung

January 19, 2024

Seoul National University

Outline

1. Introduction
2. Methodology
3. Experiment
4. Conclusion

Introduction

1. Challenge
   - In a scenario where the sensitive attribute is imbalanced, **the generalization of fairness constraints** (ex. EqualizedOdds) is substantially worse than the generalization of classification error.
2. Solution
   - FIFA: **F**lexible and **I**mbalance-**F**airness-**A**ware approach that takes both classification error and fairness constraints violation into account when training the model.

Methodology

Methodology

Notations

1. Datasets: (x, y, a)
   - $x \in \mathcal{X}$: feature vector.
   - $y \in \mathcal{Y}$: the corresponding label.
   - $a \in \mathcal{A}$: sensitive attribute
2. Task: Supervised k-class classification problem
   - model $f : \mathcal{X} \to \mathcal{R}^k$ provides k scores

$$h(x) = \underset{\mathbf{i}}{\mathbf{argmax}}\ f(x)_i$$

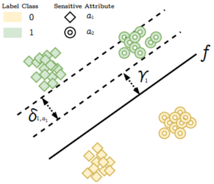
   - $h(X)$ is the prediction of the label Y of input X.
3. Classification task objective function is balanced loss.

$$\mathcal{L}_{\text{bal}}\,[f] = \mathbb{P}_{(X,Y)\sim\mathcal{P}_{\mathbf{bal}}}\left[f(X)_Y < \max_{l \neq Y} f(X)_l\right]$$

   - $\mathcal{P}_{\text{bal}} = \sum_{i=1}^{k} \mathcal{P}_i / k$ and $\mathcal{P}_i = \mathcal{P}(X \mid Y = i)$

Methodology

Notations



❺ Margin for class i by $\gamma_i = \min_{j \in S_i} \gamma(x_j, y_j)$, where

$$\gamma(x, y) = f(x)_y - \max_{l \neq y} f(x)_l$$

❻ Margin for demographic subgroups $\gamma_{i,a} = \gamma_i + \delta_{i,a}$ and $\delta_{i,a} \geq 0$

$$\gamma_i = \min\{\gamma_{i,a_1}, \gamma_{i,a_2}\}$$

Fairness Constraints

1. Violation of fairness constraints : $\mathcal{L}_{\text{fv}}$
   - In the case of binary classification & Equalized Odds

$$\mathcal{L}_{\text{fv}} = \sum_{i \in \mathcal{Y}} \left| \mathbb{P}\left(h(X) = i \mid Y = i, A = a_1\right) - \mathbb{P}\left(h(X) = i \mid Y = i, A = a_2\right) \right|$$

2. The new objective

$$\text{combined error loss: } \mathcal{M}[f] = \mathcal{L}_{\text{bal}}\left[f\right] + \alpha \mathcal{L}_{\text{fv}}$$

   - $\alpha > 0$ is hyperparameter.

Methodology

Upper bound for $\mathcal{M}[f]$ & Optimization

Theorem

With high probability, for $\mathcal{Y} = \{0, 1\}, \mathcal{A} = \{a_1, a_2\}$, and for some proper complexity measure of class $\mathcal{F}$, i.e. $C(\mathcal{F})$, for any $f \in \mathcal{F}$,

$$\mathcal{M}[f] \lesssim \sum_{i \in \mathcal{Y}} \frac{1}{\gamma_i} \sqrt{\frac{C(\mathcal{F})}{n_i}} + \sum_{i \in \mathcal{Y}, a \in \mathcal{A}} \frac{2\alpha}{\gamma_i} \sqrt{\frac{C(\mathcal{F})}{n_{i,a}}}$$

- Optimizing the upper bound in Theorem with respect to margins in the sense that

$$g(\gamma_0, \gamma_1) \leq g(\gamma_0 - \delta, \gamma_1 + \delta)$$

for $g(\gamma_0, \gamma_1) = \sum_{i \in \mathcal{Y}} \frac{1}{\gamma_i \sqrt{n_i}} + 2\alpha \sum_{i \in \mathcal{Y}, a \in \mathcal{A}} \frac{1}{\gamma_i \sqrt{n_{i,a}}}$ and all $\delta \in [-\gamma_1, \gamma_0]$

$$\gamma_0 / \gamma_1 = \tilde{n}_1^{1/4} / \tilde{n}_0^{1/4}$$

Methodology

FIFA

- Optimization

$$\gamma_0/\gamma_1 = \tilde{n}_1^{1/4}/\tilde{n}_0^{1/4},$$

where the adjusted sample size $\tilde{n}_i = \frac{n_i \Pi_{a\in\mathcal{A}} n_{i,a}}{\left(\sqrt{\Pi_{a\in\mathcal{A}} n_{i,a}} + 2\alpha \sum_{a\in\mathcal{A}} \sqrt{n_i n_{i,a}}\right)^2}$ for $i \in \{0, 1\}$.

- Since $\gamma_{i,a} = \gamma_i + \delta_{i,a}$, our target margin is as follows.

$$\gamma_{i,a} = C/\tilde{n}_i^{1/4} + \delta_{i,a}$$

where $\delta_{i,a}$ and C are non-negative parameters.

Methodology

FIFA: How to choose $\delta_{i,a}$?

- Within each class i, we identify $S_{i,a}$ with the largest subgroup $|S_{i,a}|$
  - Set the corresponding $\delta_{i,a} = 0$.
  - $\delta_{i,\mathcal{A}\setminus a} = \beta$, where $\beta \geq 0$; hyperparameter.
- As a further illustration, without loss of generality, assume for all $i, |S_{i,a_1}| \geq |S_{i,a_2}|$. Thus selected $\{\delta_{i,a}\}_{i,a}$ ensures the upper bound in the Theorem is tighter in the sense that for any $\delta > 0$,

$$\sum_{i\in\mathcal{Y}}\left(\frac{1}{\gamma_i\sqrt{n_{i,a_1}}}+\frac{1}{(\gamma_i+\delta)\sqrt{n_{i,a_2}}}\right)\leq\sum_{i\in\mathcal{Y}}\left(\frac{1}{(\gamma_i+\delta)\sqrt{n_{i,a_1}}}+\frac{1}{\gamma_i\sqrt{n_{i,a_2}}}\right).$$

FIFA: Implementation

- Consider a logits-based loss $\ell((x, y); f) = \ell\left(f(x)_y, \{f(x)_i\}_{i \in \mathcal{Y} \setminus y}\right)$
  - Ex) 0-1 loss: $1\left\{f(x)_y < \max_{i \in \mathcal{Y} \setminus y} f(x)_i\right\}$
  - Ex) Softmax-cross-entropy loss: $-\log e^{f(x)_y} / \left(e^{f(x)_y} + \sum_{i \neq y} e^{f(x)_i}\right)$.
- FIFA loss

$$\ell_{\text{FIFA}}\left((x, y, a); f\right) = \ell\left(f(x)_y - \Delta_{y,a}, \{f(x)_i\}_{i \in \mathcal{Y} \setminus y}\right)$$

where $\Delta_{i,a} = C / \tilde{n}_i^{1/4} + \delta_{i,a}$

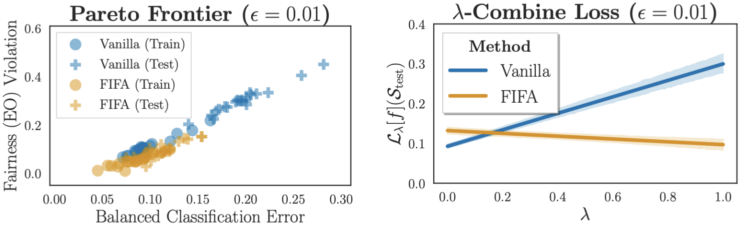
Experiment

Experiment

Method		FIFA	LDAM	Vanilla
Combined Loss	Test	**6.71%**	7.29%	14.01%
	Gen Error	**0.66%**	2.07%	6.87%
Fairness Violation	Test	**2.75%**	5.39%	20.29%
	Gen Error	**2.57%**	3.07%	13.59%
Balanced Error	Test	10.67%	9.20%	**7.74%**
	Gen Error	1.25%	1.07%	**0.15%**

- FIFA: ResNet-18 with FIFA loss for the CelebA dataset.
- LDAM: ResNet-18 using Label Distribution-Aware Margin loss(minimizing the upper bound of $\mathcal{L}_{\mathrm{bal}}$).
- Vanilla: ResNet-18 using softmax-cross-entropy loss under EO constraints.

Experiment



- Results of the 20 experiment of the balanced loss ($\mathcal{L}_{\text{bal}}$) and fairness loss ($\mathcal{L}_{\text{fv}}$) using ResNet-18 with FIFA and vanilla softmax cross-entropy loss respectively.
- λ- weighted combined loss

$$\mathcal{L}_\lambda = \lambda \mathcal{L}_{\text{bal}} + (1 - \lambda)\mathcal{L}_{\text{fv}}$$

Conclusion

Conclusion

- FIFA approach is shown to mitigate poor fairness generalization observed in vanilla models large or small.