Data Augmentation for Fairness under Spatio-Temporal Distribution Shift and Class Imbalance

S. Siamak Ghodsi
L3S-Hannover & Dept. of Math and CS, Free University Berlin

INTRODUCTION

Distribution shift: difference in training and deployment sets. Could happen due to e.g. spatial (geographical) and temporal differences in data.

- Can lead to unfairness for certain population groups.
- Gets aggravated in the presence of class imbalance.

MOTIVATION

Recently released “Folktables” dataset [1] that includes census information for the United States, provides a good example of distribution shift and imbalance. It has spatial and temporal shifts among each state’s data to the other states and has a wide range of imbalance ratio (IR) difference between states.

CASE STUDY

In this study [3] we focus on the effect of spatial distribution shifts and imbalance ratio for the year 2019. The racial distribution across the states (Fig 1) shows significant differences among the states and w.r.t. the overall US dataset. We consider each state as a context.

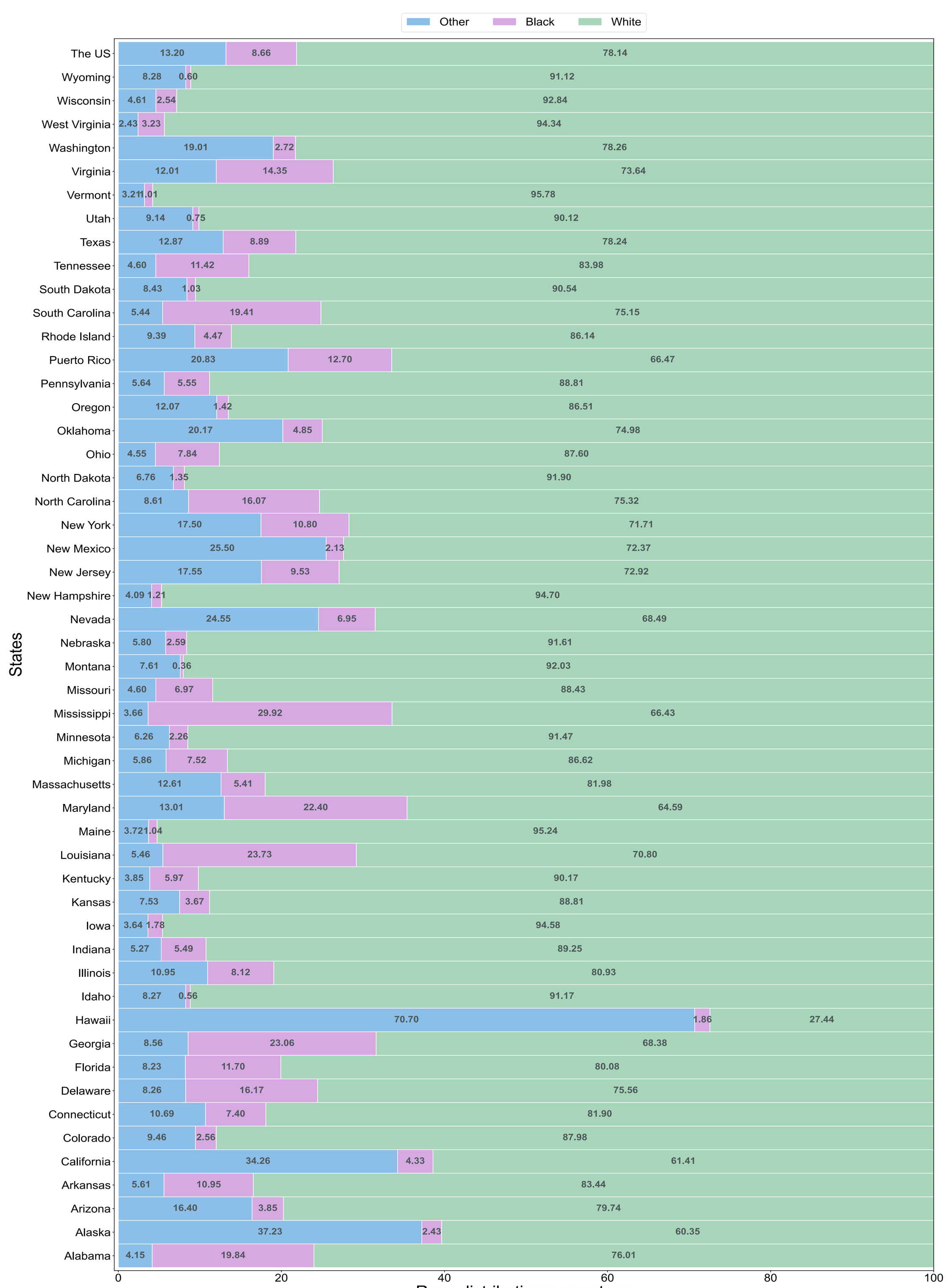

Fig 1. Percentage of racial groups (categories: {White, Black, Other}) per state, incl. US. On the 2019 dataset

RESEARCH QUESTIONS

RQ1 - Local vs Global model: *How local models learned from particular states compare to a global model trained upon data from the whole US, w.r.t both predictive and fairness-related performance?*

$$\delta FNR = \left|P(\hat{Y}=0|Y=1,g=w) - P(\hat{Y}=0|Y=1,g=b/o)\right|$$

$$\delta FPR = \left|P(\hat{Y}=1|Y=0,g=w) - P(\hat{Y}=1|Y=0,g=b/o)\right|$$

$$Eq.Odds = |\delta FPR| + |\delta FNR|$$

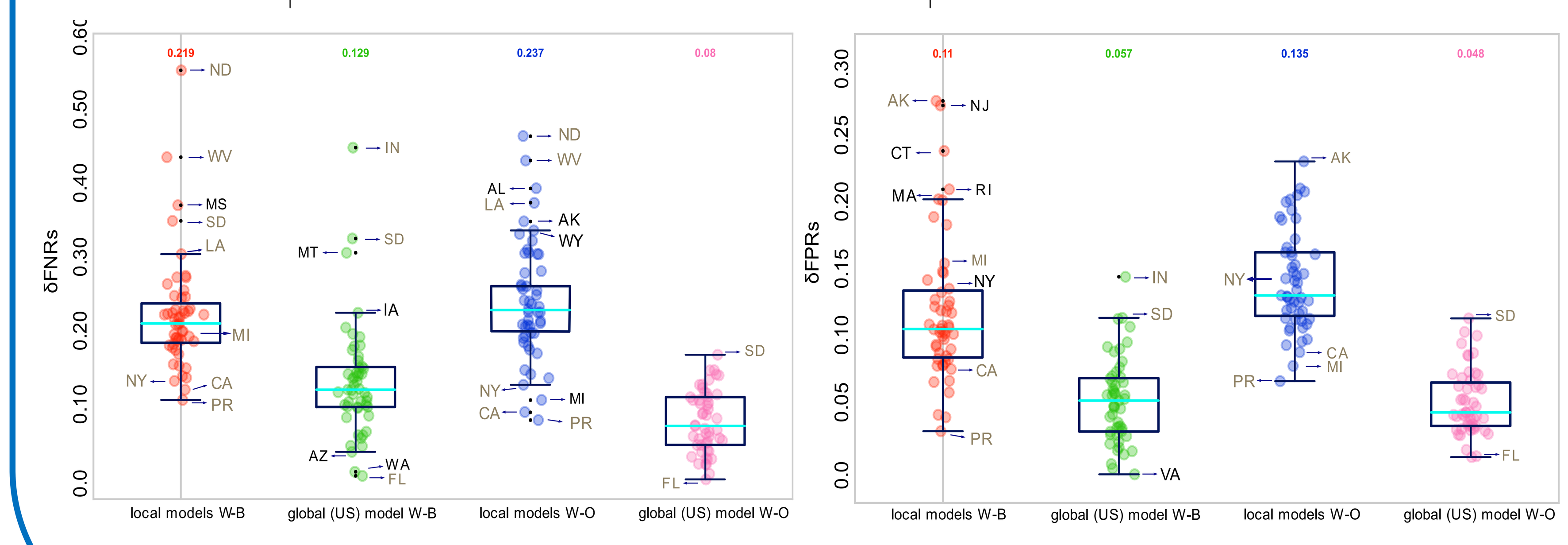

Fig 2. Spatial distribution shifts vs fairness: δFPR and δFNR scores on W-B and W-O subgroups for local/global (Logistic Regression) models.

RQ2 - Understanding spatial differences using context similarity: *How to detect context similarity, i.e., similar states, which can be used to predict how a model will perform to a different context/state, w.r.t both predictive and fairness-related performance?*

$$MMD^2(P,Q) = \|\mu_P - \mu_Q\|^2_{\mathcal{H}} \longrightarrow MMD^2(P,Q) = \left\|\frac{1}{n}\sum_{i=1}\phi(x_i) - \frac{1}{m}\sum_{i=1}\phi(v_i)\right\|^2_{\mathcal{H}}$$

$$MMD^2(X,V) = \frac{1}{m(m-1)}\sum_i\sum_{j\neq i}k(\mathbf{x_i},\mathbf{x_j}) - 2\frac{1}{m.m}\sum_i\sum_j k(\mathbf{x_i},\mathbf{v_j}) + \frac{1}{m(m-1)}\sum_i\sum_{j\neq i}k(\mathbf{v_i},\mathbf{v_j})$$

Context similarity using Maximum mean discrepancy **(MMD)** [2]: calculates distances between datasets (*X,V*) as distances between their probability distributions (*P,Q*). kernel trick: calculate distance between latent feature mean embeddings. MMD only uses feature-values (not classes)

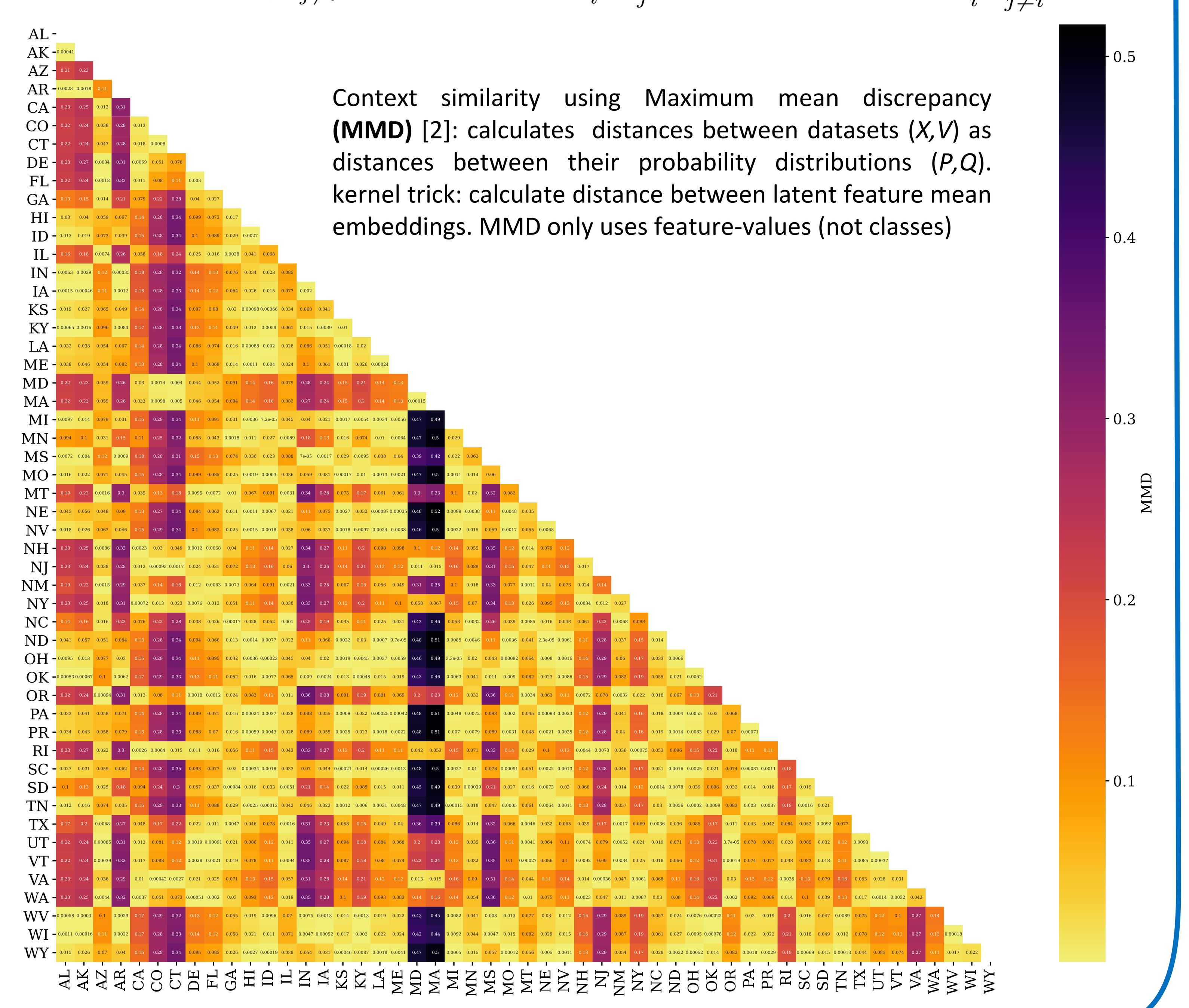

Fig 3. Local to Local Similarity: Pairwise local context similarity (MMD) between states. Light colors indicate higher similarity, dark stand for smaller similarity

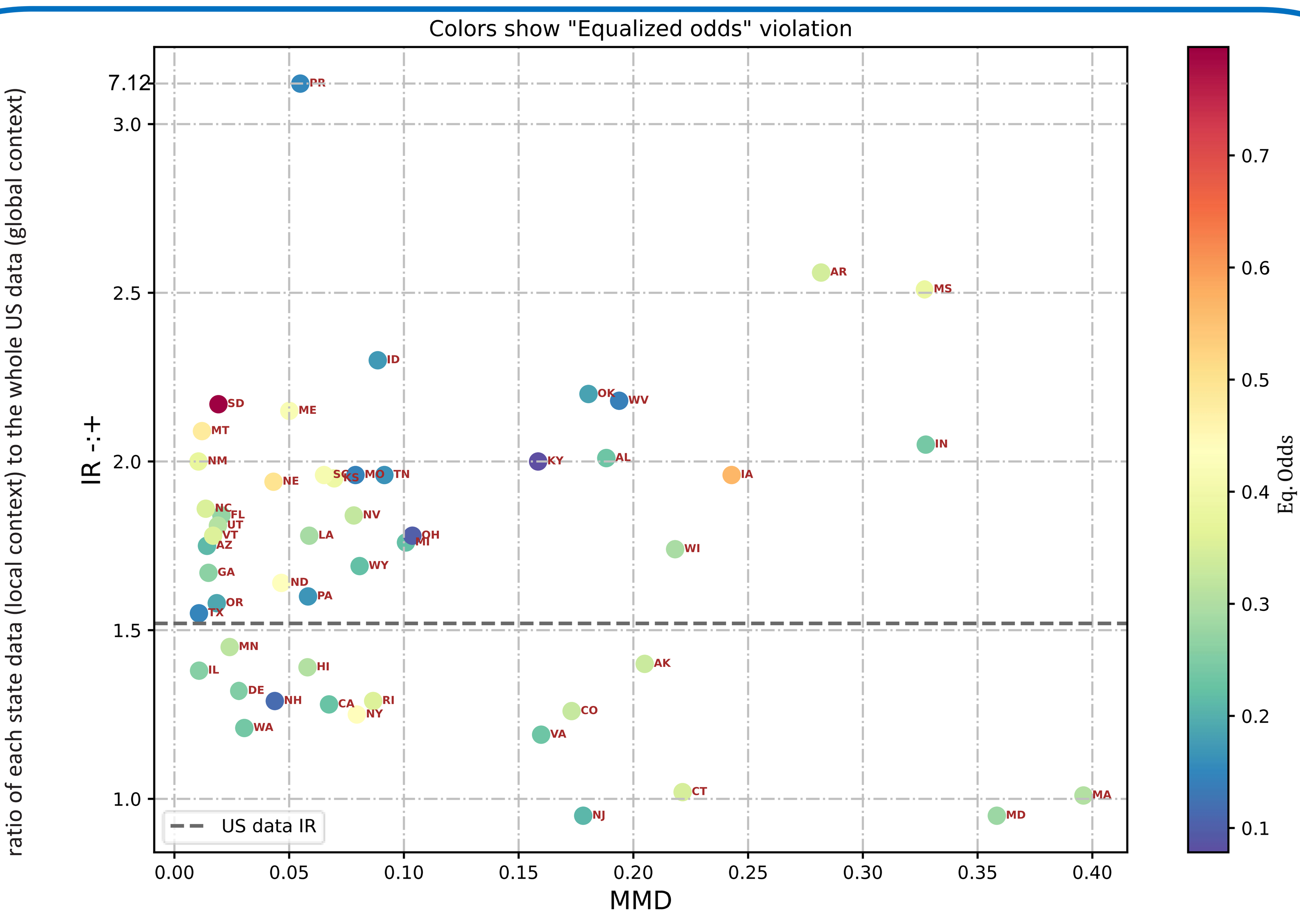

Fig 4. Local to Global Similarity: comparing data distribution and imbalance ratio of each state data (local context) to the whole US data (global context)

DISCUSSION ON RESULTS

In Fig 2. global model (green and pink boxplots) seems generally to be less discriminative in deployment, than the local models. However, it doesn’t perform same for different context. It performs better on states that have more similar data distribution to it than the ones less similar. For example, for *VA* compared to *IN* or *FL* compared to *SD*. Looking at Fig 1 (FL has much more similar racial group distribution to the US data compared to SD). Same for *VA and IN*. In Fig 4, MMD value close to 0 (means very similar) and also being closer to 1.5 IR value (dashed line) shows similarity to the US data (*VA and FL* much closer than *IN and SD* respectively). In Fig 3, MA has worst similarity to other contexts (states) correlates to red δFPR boxplot. NY and CA show much better similarity to the other states (lighter color) in Fig 3, and perform as the best local models in Fig 2.

CHALLENGES AND FUTURE DIRECTIONS

Challenges: As seen in boxplots, a global model performs less discriminatively compared to local models, but still it doesn’t perform fairly/similarly on all the states. So, a problem is high variance in deployment discrimination-score on different states (**unreliability of global model**). Another problem is estimating an application range for local models (**clustering similar context that perform similarly**). But how is similarity defined? Spatial neighbors (**geopolitical similarity**) can be similar, **semantically similar contexts** (based on a similarity score e.g. **MMD**) can be also similar.

Future Direction: building an augmented fair local-model using the similarity notion that outperforms each single local model and is comparable or even better (less discriminative) than the global model.

Question: Model augmentation using Similar Context vs Synthetic data generation. Which will perform better?

REFERENCES & CONTRIBUTIONS

1. Ding F., Hardt M., Miller J., Schmidt L.: Retiring adult: New datasets for fair machine learning. In: NeurIPS. pp. 6478–6490 (2021).
2. Gretton A., Borgwardt K.M., Rasch M.J., Scholkopf B., Smola A.J.: A kernel two-sample test. J. Mach. Learn. Res. 13, 723–773 (2012).
3. **Ghodsi Siamak**, Harith Alani, and Eirini Ntoutsi. "Context matters for fairness - a case study on the effect of spatial distribution shifts." arXiv preprint arXiv:2206.11436 (2022). **Submitted, Pre-print available**
4. **Ghodsi Siamak**, Harith Alani, and Eirini Ntoutsi. “A context-aware fair learning model using local similarity for augmentation." **Under progress**
5. **Ghodsi Siamak**, Vasilis Iosifidis, Arjun Roy, and Eirini Ntoutsi. “Fair-SMOTEBoost: Sub-group correction for parity-based cumulative fairness-aware boosting.“ **Under progress**

Find more about me:

- Webpage: https://siamakghodsi.github.io/
- ghodsi@{l3s.de, zedat.fu-berlin.de}
- @SiamakGhodsi

SCAN ME
Watch this poster online

NÔBIAS

This work has received funding from the European Union’s Horizon 2020 research and innovation programme under Marie Sklodowska-Curie Actions (grant agreement number 860630) for the project “NoBIAS - Artificial Intelligence without Bias”.