Fair Risk Predictions for Underrepresented Populations Using Electronic Health Records

Judy Zhong

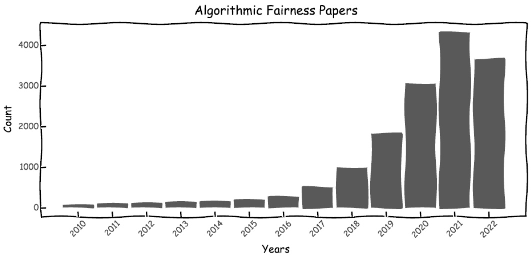
NYU Grossman School of Medicine

Division of Biostatistics Seminar

Oct 31, 2022

Introduction

- In recent years, the machine learning community has become alert to the ways that predictive models can introduce unfairness in decision-making.
- Unfairness is defined as the disparity in prediction performances between subgroups[1].
- Examples include: recidivism prediction, credit worthiness, facial recognition, job recommendation/listing, ...
- To address this issue, there has been a significant body of work in the machine learning community on algorithmic fairness.



[1] e.g. race/ethnicity, gender

Introduction

- The healthcare community also became alert to this problem (Mhasawade et al., 2021; Fuster et al., 2022; Xu et al., 2022). For example, we found substantial disparities in the Electronic Health Record (EHRs):
  - Minority patients and patients with disadvantaged social determinants of health are often under-represented in terms of sample sizes, number of encounters, and number of lab results (predictors)
- Despite all these efforts, most of the existing work has focused on predictions for binary classification.
- Thus, there is a gap between the practical use of models for various types of outcomes (e.g. count data) and the development of fairness-aware methodologies for those models.
- In this study, we develop a framework **to achieve fair predictions**.

Introduction

Do the prediction performance disparities exist?

.. if we apply generalized linear models?

Introduction

Motivating Example: Performance Disparity of GLMs on Benchmark Datasets

Outcome Type	Dataset	Group (K)	Test MSE	Disparity[a]	Rel. Disp.(Disp./MSE)
Binary	Adult	Gender (2)	0.105	0.023	21%
	Arrhythmia	Gender (2)	0.258	0.029	11%
	COMPAS	Race (4)	0.211	0.004	2%
	Drug	Race (2)	0.112	0.092	20%
	German	Gender (2)	0.183	0.045	25%
Continuous	Crime	Race (3)	0.057	0.092	**163%**
	Law School	Race (5)	0.840	0.148	18%
	Parkinsons	Gender (2)	93.112	31.416	34%
	Student	Gender (2)	0.677	0.203	30%
Count	HRS	Race (4)	0.579	0.300	**52%**
Multiclass	Drug	Race (2)	0.079	0.005	7%
	HCV	Gender (2)	0.020	0.015	**76%**
	Obesity	Gender (2)	0.062	0.032	**51%**

[a] average absolute difference of groupwise MSEs

Objective

- Goal: Develop a fair Generalized Linear Model to reduce the prediction performance disparity between subgroups, while not decreasing the overall performance as much as possible

Demographic Parity

- Suppose we are given K groups defined by a *sensitive attribute* A (e.g. race/ethnicity, gender, or such).

Definition: Demographic Parity (Kamiran and Calders, 2009; Hardt et al., 2016)

A GLM satisfies *demographic parity* (DP) if its prediction $f(\mathbf{X})$ is statistically independent of the sensitive attribute A. That is, $\mu(\mathbf{X}\beta) \perp\!\!\!\perp A$.

Fair GLMs

Thus, the MMD fairness penalty term is:

$$\mathcal{D}(\beta; \mathbf{X}, \mathbf{a}) = \sum_{k<l} \mathrm{MMD}^2(\mathbf{X}_k\beta, \mathbf{X}_l\beta). \tag{1}$$

$$\widehat{\beta}_{\mathrm{FGLM}} = \underset{\beta}{\mathrm{argmin}} \; -\sum_{k=1}^{K} \sum_{i=1}^{n_k} \ell(\beta; \mathbf{x}_{ki}, y_{ki}) + \lambda \mathcal{D}(\beta; \mathbf{X}, \mathbf{a}). \tag{2}$$

Case Study

Motivating Example Revisited

Outcome	Dataset	GLM MSE	GLM Rel. Disp.	FGLM-LMMD MSE	FGLM-LMMD Rel. Disp.	FGLM-GMMD MSE	FGLM-GMMD Rel. Disp.
Binary	Arrhythmia	0.27	11%	0.28 (▲0.01)	10% (▼1%p)	0.27 (▲0.01)	0% (▼11%p)
	COMPAS	0.21	2%	0.22 (▲0.01)	1% (▼1%p)	0.22 (▲0.01)	1% (▼1%p)
	Drug	0.11	15%	0.11 (0.00)	17% (▲2%p)	0.11 (0.00)	12% (▼3%p)
	German	0.18	24%	0.19 (▲0.01)	21% (▼3%p)	0.18 (0.00)	21% (▼3%p)
Continuous	Crime	0.06	148%	0.04 (▼0.02)	72% (▼76%p)	0.04 (▼0.02)	70% (▼78%p)
	Parkinsons	97.8	35%	92.6 (▼5.2)	34% (▼1%p)	95.8 (▼2.0)	27% (▼7%p)
	Student	0.67	30%	0.67 (0.00)	29% (▼1%p)	0.67 (0.00)	29% (▼1%p)
Count	HRS	0.58	52%	0.59 (▲0.01)	50% (▼2%p)	0.58 (0.00)	47% (▼5%p)
Multiclass	Drug	0.08	4%	0.08 (0.00)	2% (▼2%p)	0.08 (0.00)	0% (▼4%p)
	HCV	0.02	41%	0.02 (0.00)	32% (▼9%p)	0.02 (0.00)	5% (▼36%p)
	Obesity	0.06	36%	0.06 (0.00)	35% (▼1%p)	0.07 (▲0.01)	20% (▲16%p)

*Hyperparameters of the fair models are selected for their MSEs to remain below 110% of GLM's MSE

Conclusion

- The naive method may generate disparate prediction performances for the under-represented subpopulations
- The proposed fair model can effectively reduce prediction disparity while maintaining the overall prediction performances
- It is applicable to most types of outcomes

Future Directions

- **Parent R01 Project**: Apply the fair GLMs to improve prediction fairness for the under-represented sub-populations in the presence of unbalanced sample sizes and covariates
- **Methodological Directions** Improve the fair GLMs with mis-labeled sensitive attributes and missing data, which often are found in EHRs
- **Time-to-event Models** Fairness-aware survival analysis methods have gotten less attention so far. A similar approach could be applied to survival analysis models.

References and Publications

Do, H., Nandi, S., Putzel, P., Smyth, P., and Zhong, J. (2022a). A joint fairness model with applications to risk predictions for underrepresented populations. *Biometrics*.

Do, H., Putzel, P., Martin, A. S., Smyth, P., and Zhong, J. (2022b). Fair generalized linear models with a convex penalty. In *International Conference on Machine Learning*, pages 5286–5308. PMLR.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., and Walther, A. (2022). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, 77(1):5–47.

Hardt, M., Price, E., Price, E., and Srebro, N. (2016). Equality of opportunity in supervised learning. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

Kamiran, F. and Calders, T. (2009). Classifying without discriminating. In *2009 2nd International Conference on Computer, Control and Communication*, pages 1–6.

Mhasawade, V., Zhao, Y., and Chunara, R. (2021). Machine learning and algorithmic fairness in public and population health. *Nature Machine Intelligence*, 3(8):659–666.

Xu, J., Xiao, Y., Wang, W. H., Ning, Y., Shenkman, E. A., Bian, J., and Wang, F. (2022). Algorithmic fairness in computational medicine. *eBioMedicine*, 84:104250.