Fairness with Censorship: Bridging the Gap between Fairness Research and Real-World Deployment

Wenbin Zhang

Florida International University, FL, USA
wenbin.zhang@fiu.edu

Talk Summary

Real-life instances of Artificial Intelligence (AI)-based decision-making systems have consistently demonstrated discrimination and bias, particularly towards marginalized groups. Consequently, a burgeoning body of research has emerged to quantify and mitigate unfairness in AI. Most of them address fairness by assuming the presence of the class label, in which the fairness notions are defined based on the class label either actual or predicted, and the same predictive model is trained contingent upon for new instance prediction. However, limited attention has been given to censoring settings, where the class label could be absent due to censorship on the time to an event of interest, hindering the application of existing fairness notions and approaches.

On the other hand, the censoring phenomenon is widespread in various real-world applications and fairness benchmark datasets, such as predicting criminal recidivism (COMPAS dataset), customer profiling for business planning (KKBOX dataset), and estimating the probability of a specific disease or clinical outcome in clinical prediction (METABRIC dataset), among many others, in which the participants could lost to follow-up, withdraw from the study or experiencing a competing event, etc., leading to the absent of class label. Due to the inability to handle censorship information, existing fairness studies focus on bias quantification and mitigation of class label proportions by either excluding observations with uncertain class labels due to censorship or omitting censorship information for these instances. However, both of them contain crucial information, and their removal could bias results even towards those with known class labels. To this end, this talk revisits fairness and reveals idiosyncrasies of existing fairness literature assuming the availability of class label that limits their real-world utility. The primary artifacts are formulating fairness with censorship to account for scenarios where the class label is not guaranteed, and a suite of corresponding new fairness notions, algorithms, and theoretical constructs (Zhang and Weiss 2021, 2022, 2023; Zhang, Hernandez-Boussard, and Weiss 2023; Zhang et al. 2023). I argue that this formulation has a broader applicability to practical scenarios concerning fairness. I also show how the newly devised fairness notions involving censored information and the general framework for fair predictions in the presence of censorship allow us to measure and mitigate discrimination amidst censorship that bridges the gap between the design of a "fair" model in the lab and its deployment in the real-world.

Biography

Dr. Wenbin Zhang is an Assistant Professor in the Knight Foundation School of Computing & Information Sciences at Florida International University, and an Associate Member at the Te Ipu o te Mahara Artificial Intelligence Institute. His research investigates the theoretical foundations of machine learning with a focus on societal impact and welfare. In addition, he has worked in a number of application areas, highlighted by work on healthcare, geophysics, transportation, forestry, and finance. He is a recipient of the NSF CRII Award, and best paper awards/candidates at FAccT'23, ICDM'23, DAMI and ICDM'21. He also regularly serves in the organizing committees across computer science and interdisciplinary venues, most recently Travel Award Chair at AAAI'24, Volunteer Chair at WSDM'24, and Student Program Chair at AIES'23.

Acknowledgments

This work was supported in part by National Science Foundation (NSF) under Grant No. 2245895.

References

Zhang, W.; Hernandez-Boussard, T.; and Weiss, J. 2023. Censored fairness through awareness. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 14611–14619.

Zhang, W.; Wang, Z.; Kim, J.; Cheng, C.; Oommen, T.; Ravikumar, P.; and Weiss, J. 2023. Individual Fairness under Uncertainty. In *ECAI 2020*, 3042–3049.

Zhang, W.; and Weiss, J. 2021. Fair Decision-making Under Uncertainty. In *2021 IEEE International Conference on Data Mining (ICDM)*. IEEE.

Zhang, W.; and Weiss, J. C. 2022. Longitudinal fairness with censorship. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 12235–12243.

Zhang, W.; and Weiss, J. C. 2023. Fairness with censorship and group constraints. *Knowledge and Information Systems*, 1–24.