

Improving Fairness Generalization Through a Sample-Robust Optimization Method

Ulrich Aïvodj¹, Julien Ferry², Sébastien Gambs¹,
Marie-José Huguet² and Mohamed Siala²

¹ École de Technologie Supérieure, Montréal, Canada

² LAAS-CNRS, Université de Toulouse, CNRS, INSA, Toulouse, France
jferry@laas.fr

November 2022

Summary

In this talk, we present a novel approach for improving fairness generalisation in predictive decision making. The work is published recently in the Machine Learning journal [1]. The expected duration of the talk is about 15 minutes.

Abstract

Unwanted bias is a major concern in machine learning, raising in particular significant ethical issues when machine learning models are deployed within high-stakes decision systems. A common solution to mitigate it is to integrate and optimize a statistical fairness metric along with accuracy during the training phase. However, one of the main remaining challenges is that current approaches usually generalize poorly in terms of fairness on unseen data. We address this issue by proposing a new robustness framework for statistical fairness in machine learning. The proposed approach is inspired by the domain of Distributionally Robust Optimization and works in ensuring fairness over a variety of samplings of the training set. Our approach can be used to quantify the robustness of fairness but also to improve it when training a model. We empirically evaluate the proposed method and show that it effectively improves fairness generalization. In addition, we propose a simple yet powerful heuristic application of our framework that can be integrated into a wide range of existing fair classification techniques to enhance fairness generalization. Our extensive empirical study using two existing fair classification methods demonstrates the efficiency and scalability of the proposed heuristic approach.

Keywords: Fairness in Machine Learning, Combinatorial Optimisation

References

- [1] Julien Ferry, Ulrich Aïvodji, Sébastien Gambs, Marie-José Huguet, and Mohamed Siala. Improving fairness generalization through a sample-robust optimization method. *Machine Learning*, Jul 2022.