Towards Reliable and Practicable Algorithmic Recourse

Himabindu Lakkaraju
Harvard University
hlakkaraju@seas.harvard.edu

ABSTRACT

As predictive models are increasingly being deployed in high-stakes decision making (e.g., loan approvals), there has been growing interest in developing post hoc techniques which provide recourse to individuals who have been adversely impacted by predicted outcomes. For example, when an individual is denied loan by a predictive model deployed by a bank, they should be informed about reasons for this decision and what can be done to reverse it. While several approaches have been proposed to tackle the problem of generating recourses, these techniques rely heavily on various restrictive assumptions. For instance, these techniques generate recourses under the assumption that the underlying predictive models do not change. In practice, however, models are often updated for a variety of reasons including data distribution shifts. There is little to no research that systematically investigates and addresses these limitations.

In this talk, I will discuss some of our recent work that sheds light on and addresses the aforementioned challenges, thereby paving the way for making algorithmic recourse practicable and reliable. First, I will present theoretical and empirical results which demonstrate that the recourses generated by state-of-the-art approaches are often invalidated due to model updates. Next, I will introduce a novel algorithmic framework based on adversarial training to generate recourses that remain valid even if the underlying models are updated. I will conclude the talk by presenting theoretical and empirical evidence for the efficacy of our solutions, and also discussing other open problems in the burgeoning field of algorithmic recourse.

CCS Concepts/ACM Classifiers

• Computing methodologies ~ Machine learning

Author Keywords

Explainability; Algorithmic Recourse;

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author(s).
CIKM '21, November 1–5, 2021, Virtual Event, Australia.
© 2021 Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-8446-9/21/11.
https://doi.org/10.1145/3459637.3482497

BIOGRAPHY

Himabindu (Hima) Lakkaraju is an assistant professor at Harvard University focusing on explainability, fairness, and robustness of machine learning models. She has also been working with various domain experts in policy and healthcare to understand the real-world implications of explainable and fair ML. Hima has been named as one of the world's top innovators under 35 by both MIT Tech Review and Vanity Fair. Her research has also received best paper awards at SIAM International Conference on Data Mining (SDM) and INFORMS, and grants from NSF, Google, Amazon, and Bayer. Hima has given keynote talks at various workshops at ICML, NeurIPS, AAAI, and CVPR, and her research has also been showcased by popular media outlets including the New York Times, MIT Tech Review, TIME magazine, and Forbes. More recently, she co-founded the Trustworthy ML Initiative to enable easy access to resources on trustworthy ML and to build a community of researchers/practitioners working on the topic.