Learning-Augmented Robust Algorithmic Recourse

Kshitij Kayastha Vasilis Gkatzelis Shahin Jabbari

Drexel University

Algorithmic Recourse

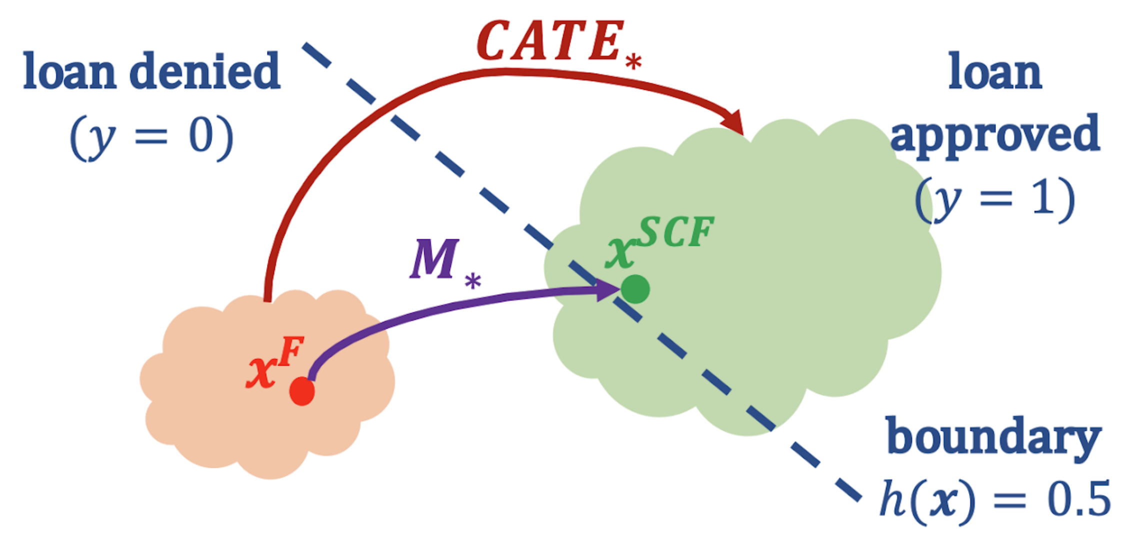


Figure 1. An example of algorithmic recourse. Image Credit: Karimi et al. [1]

- Machine learning models are nowadays widely deployed even in sensitive domains such as lending
- Recent work in responsible computing aims to make these models fair, transparent, and explainable
- **Algorithmic recourse provides individuals with undesirable labels (e.g., loan request was denied) with a minimum cost improvement suggestion to achieve the desired label**

Robustness in Recourse

- In practice, many models are periodically updated to reflect the changes in data, causing the recourse to become invalid, i.e., following it may not lead to a desirable outcome
- Previous work studied how to provide robust recourse [2] for adversarial model change, though this comes at the price of a higher cost for implementing recourse
- **How can we lower this cost while still maintaining worst-case performance guarantees?**

Learning-Augmented Framework

- The learning-augmented framework overcomes the limitations of adversarial (i.e., worst-case) analysis
- **Assumes the designer can at least formulate some predictions regarding** these changes

Problem Formulation

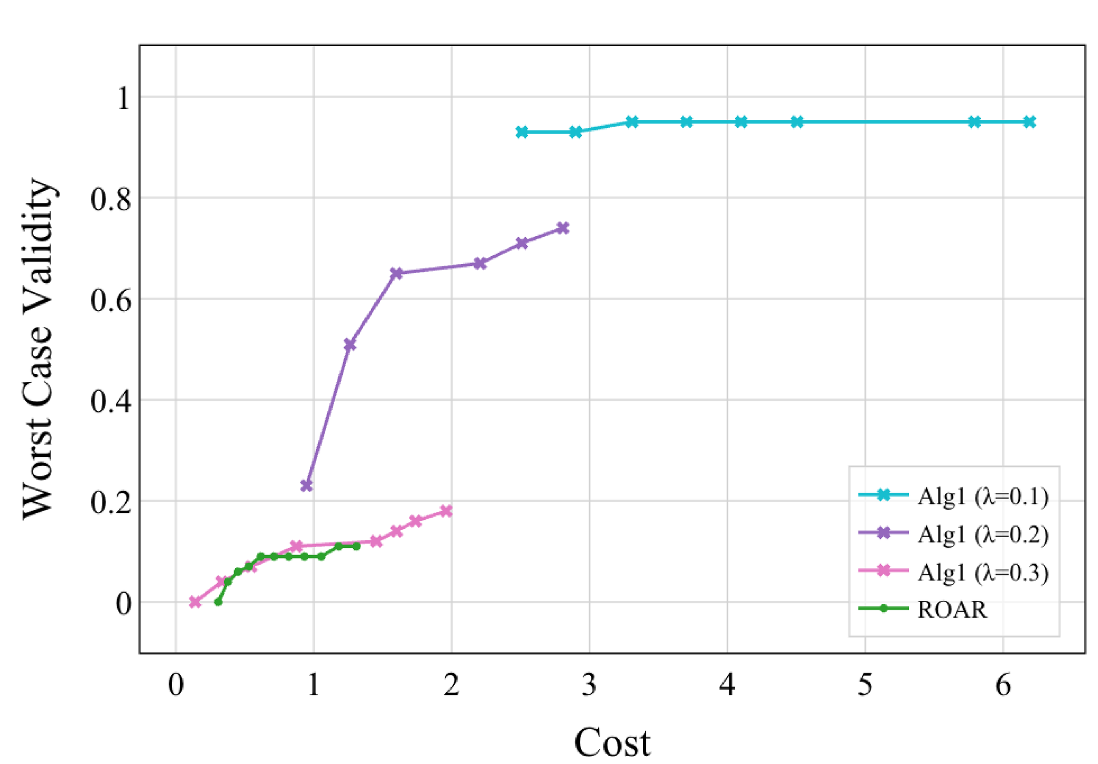
- Given access to a possibly inaccurate prediction regarding a future model, can we design an algorithm that simultaneously satisfies the following two properties:
  1. [Consistency] Compute recourses that perform near-optimally when the predictions are accurate
  2. [Robustness] Maintain good performance even in the worst-case, i.e., even when the predictions are arbitrarily inaccurate

Contributions

- For a specific class of recourse formulations, model classes, and model changes, we provide the first **computationally efficient algorithm to compute optimal robust recourse**
- We extend the learning-augmented framework to algorithmic by defining notations of consistency and robustness for algorithmic recourse
- Provide **an efficient algorithm that computes a trade-off between robustness and consistency**

Comparison of Robustness to Prior Work

- A performance of recourse can be broken down into two parts:
  1. Cost of implementing the recourse
  2. The validity of recourse in achieving the desired outcome
- Figure 2 shows our algorithm can achieve results with very high validity, albeit at 2- 3x higher implementation cost
- Prior work [2] cannot guarantee high validity



Robustness Consistency Trade-off

- Our algorithm can efficiently compute the trade-off between robustness and consistency
- Figure 3 shows that our algorithm finds a lower-cost trade-off compared to prior work

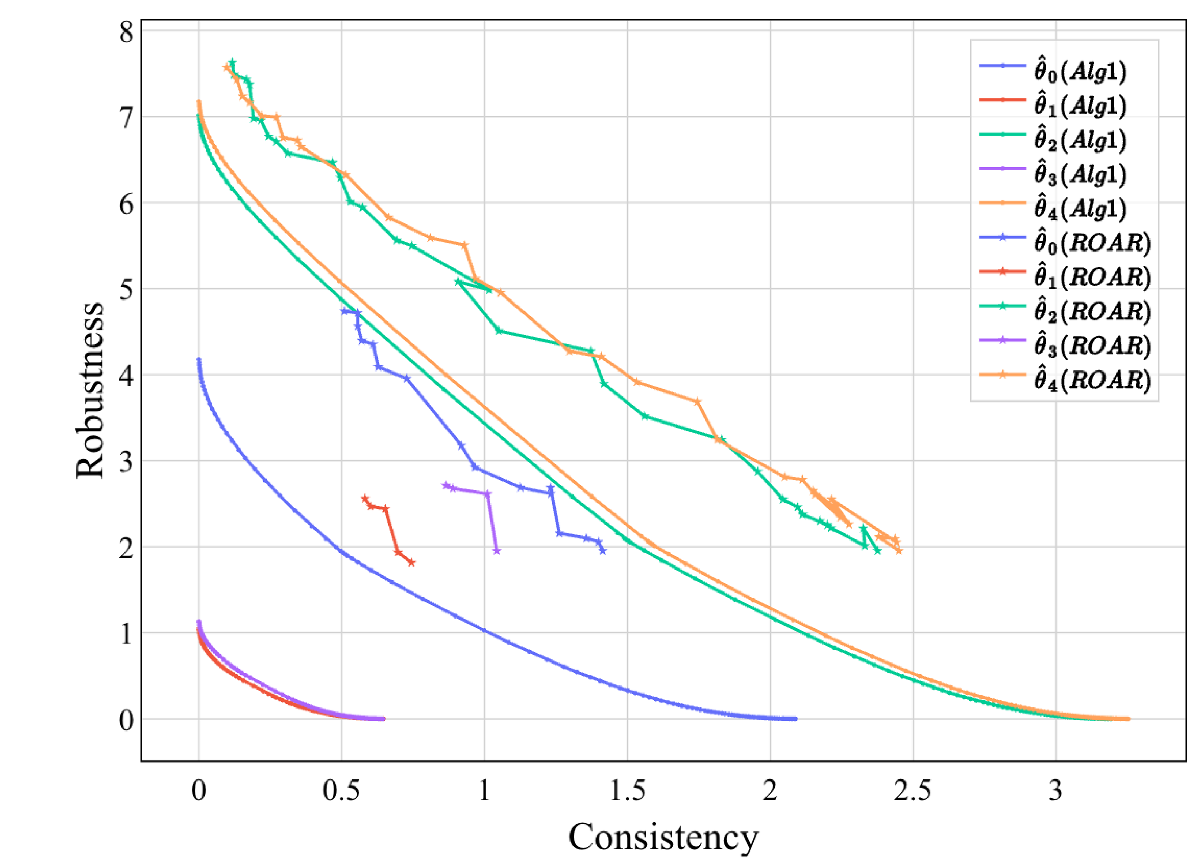


Figure 3. The trade-off between robustness and consistency of our algorithm and ROAR [2]. Smaller values are more desirable for both. Each curve corresponds to a different prediction.

Future Work

- Theoretical understanding of the Pareto frontier of the trade-off between consistency and robustness
- Theoretical understanding of how prediction error affects consistency and robustness
- Extending our framework to alternative recourse formulation, model classes, and types of model change

References

[1] Amir-Hossein Karimi, Bodo Julius von Kügelgen, Bernhard Schölkopf, and Isabel Valera. Algorithmic recourse under imperfect causal knowledge: a probabilistic approach. In *Advances in Neural Information Processing Systems 33*, 2020.

[2] Sohini Upadhyay, Shalmali Joshi, and Himabindu Lakkaraju. Towards robust and reliable algorithmic recourse. In *Advances in Neural Information Processing Systems 34*, pages 16926–16937, 2021.

https://https://shahin-jabbari.github.io/ Department of Computer Science, Drexel University, Philadelphia, PA shahin@drexel.edu