Enhancing Fairness through Time-Aware Recourse: A Pathway to Realistic Algorithmic Recommendations*

Isacco Beretta[1,†], Martina Cinquini[1,*,†] and Isabel Valera[2]

[1]Department of Computer Science, University of Pisa, Pisa

[2]Department of Computer Science, Saarland University, Saarbrücken, Germany

Abstract

Algorithmic Recourse (AR) addresses adverse outcomes in automated decision-making by offering actionable recommendations. However, current state-of-the-art methods overlook the interdependence of features and do not consider the temporal dimension. To fill this gap, TIME-CAR emerges as a pioneering approach that integrates temporal information. Building upon this formulation, this work investigates the context of fairness, specifically focusing on the implications for marginalized demographic groups. Since long wait times can significantly impact communities' financial, educational, and personal lives, exploring how time-related factors affect the fair treatment of these groups is crucial to suggest potential solutions to reduce the negative effects on minority populations. Our findings set the stage for more equitable AR techniques sensitive to individual needs, ultimately fostering fairer suggestions.

Keywords

Algorithmic Recourse, Fairness, Consequential Recommendations

1. Introduction

Algorithmic Recourse (AR) seeks to provide actionable recommendations that should be performed to reverse negative outcomes from automated decision-making systems. Recently, this field has emerged as one of the most promising solutions to explainability in Machine Learning due to its compliance with legal requirements [1], its psychological benefit for the individual [2], and its potential to explore "what-if" scenarios [3]. Among current literature, recent work [4] highlights that a significant drawback of AR methods is the implicit assumption of examining features as independently manipulable inputs. Since the individual's attributes change may have downstream effects on other features, observing and identifying causal mechanisms is crucial in analyzing real-world scenarios to avoid sub-optimal or infeasible actions. From this perspective, [5, 6] propose a fundamental reformulation of the recourse problem, incorporating knowledge of causal dependencies into recommending recourse actions. The ability to assess the causal relationships explicitly guarantees plausible counterfactuals [7] and improves the

EWAF'24: European Workshop on Algorithmic Fairness, July 01–03, 2024, Mainz, Germany

*Corresponding author.

†These authors contributed equally.

isacco.beretta@phd.unipi.it (I. Beretta); martina.cinquini@phd.unipi.it (M. Cinquini); ivalera@cs.uni-saarland.de (I. Valera)

https://marti5ini.github.io/ (M. Cinquini); https://ivaleram.github.io (I. Valera)

0000-0002-0877-7063 (I. Beretta); 0000-0002-0877-7063 (M. Cinquini); 0000-0002-6440-4376 (I. Valera)

user's perception of a decision's quality since it reflects the tendency of human beings to think in terms of cause-effect [8].

A significant limitation of current methods is their inability to incorporate the temporal dimension. Neglecting the temporal interdependencies between features and actions can result in erroneous identification of the more effective features cost and time-wise. As a result, there is a need to devise Causal AR techniques that can incorporate temporal information to provide explanations that precisely reflect the complex dynamics of the system and to guarantee that the recommendations offered are reliable and plausible.

In [9], the authors discuss the necessity of interpreting the causal model as a representation of a dynamical process that involves the evolution of its instances over time. Specifically, they introduce TIME-CAR, one of the first proposals on integrating the temporal dimension into a Causal AR problem by including the topological information of the causal graph in the cost function evaluation.

This research investigates the implications of fairness within the TIME-CAR framework, focusing on how longer periods needed for certain tasks affect marginalized demographic groups and their connection to socioeconomic stability, educational opportunities, and overall well-being. The increased time required for these tasks can intensify existing inequalities and vulnerabilities, leading to a continuous cycle of disadvantage that is hard to break. This work aims to formulate fairer AR methods sensitive to these populations' unique needs and time constraints.

2. Fairness through Time-Aware Recourse

Actionable Recourse. The problem of AR can be formulated as a constrained optimization in the following terms: given a binary classification model $h : \mathbf{X} \to \{0, 1\}$, and an instance X for which $h(X) = 0$, the goal is to determine the action $\mathbb{A}_{\delta^*}$ satisfying

$$\mathbb{A}_{\delta^*} = \arg\min_{\mathbb{A}_\delta} c(X, \mathbb{A}_\delta) \quad s.t. \quad h(\mathbb{A}_\delta(X)) = 1$$

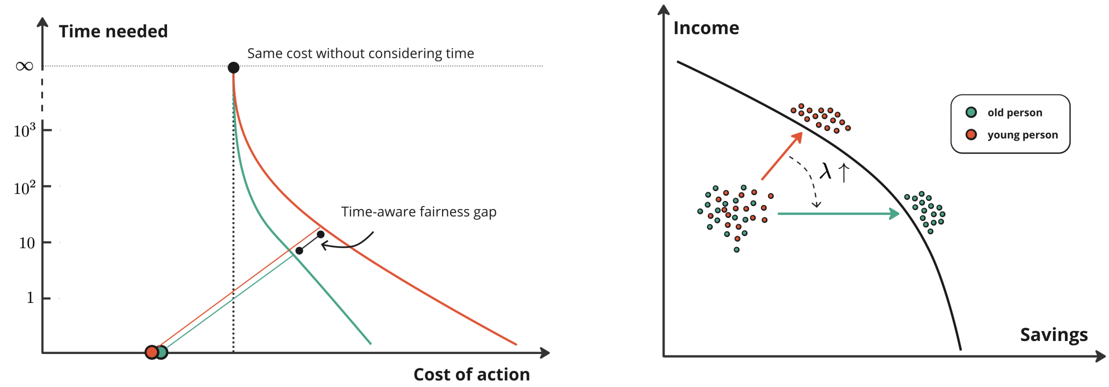
where $\mathbb{A}_\delta(X)$ represents a modified version of X. In other words, the objective is to identify the minimal cost action that alters the model's decision from unfavorable to favorable.

2.1. Same Cost, Different Times

In [9], the authors introduce a new definition of the cost of an action that incorporates the temporal dimension:

$$c(X, Y, \mathbb{A}_\delta) = c_s\left(X, \mathbb{A}_\delta\right) + \lambda c_t\left(X, \mathbb{A}_\delta, Y\right),$$

where X denotes the individual's initial state, Y is the target state (e.g., the one that guarantees loan acceptance), and $\mathbb{A}_\delta$ is the action taken to obtain the transition between them. λ is a tunable parameter that values how important is time compared to the other features. In particular, it balances the two components of the cost function, namely c_s, which denotes the cost function in the feature space, and c_t, which reflects the time part. A time-unaware recommendation algorithm is basically one that fixes $\lambda = 0$.



(a) Not accounting for time could introduce hidden biases in recommendation algorithms.

(b) Let $cost(\delta) = c_s(\delta) + \lambda \mathbb{1}(\delta_{income} \neq 0)$, it could be that λ is depending on age.

Figure 1: Possible scenarios where the time cost matters from a fairness perspective.

This section explores a scenario where sensitive attributes are included among the features, denoted as $A \subset X$. We examine the case of two individuals, i_1 and i_2, with different sensitive attributes' values, such that $A(i_1) \neq A(i_2)$. We hypothesize that the cost recommendations from the time-unaware automatic decision system for these individuals, $c_s(X(i_1), \mathbb{A}_{\delta_1})$ and $c_s(X(i_2), \mathbb{A}_{\delta_2})$, are approximately equal. This implies that despite the difference in sensitive attributes, the system suggests similar cost interventions for both individuals. However, the temporal cost c_t could vary significantly between them, meaning one individual might need more time to achieve the desired state than the other. This scenario is illustrated in Figure 1a.

2.2. Not Everyone Values Time Equally

In another scenario, time may be regarded as a resource whose value varies based on individual characteristics. Figure 1b demonstrates this idea through a specific case related to applying for a loan. The value of λ might be higher for the older population as they are likely closer to retirement and have a limited window to recuperate from financial setbacks. Conversely, younger individuals might have a lower λ value given their longer time horizon to adjust their savings behavior. Hence, financial models must be calibrated to accommodate varying λ values across different demographic segments. This understanding enables the creation of customized recommendations sensitive to each individual's dynamics and the time-related evaluation of changes within their specific societal and economic contexts.

2.3. Actionability as a Time-Constraint

In the context of AR, *plausibility* refers to the perceived consistency and reasonableness of the recommendations provided by recourse approaches. From a psychological perspective, providing plausible explanations enables users to form mental models that align with their prior knowledge [10]. When the temporal dimension is incorporated into causal reasoning, an AR approach could ensure that the actions suggested are psychologically congruent with human

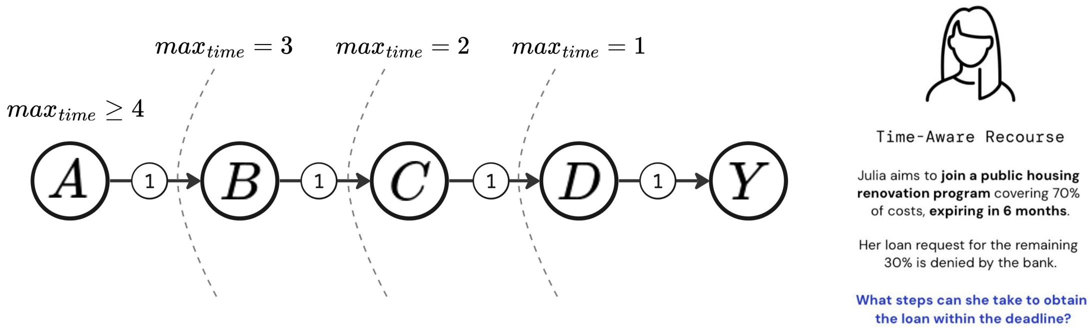


Figure 2: (Left) As max_{time} decreases, the graph undergoes a pruning process that reduces the number of actionable variables. For example, only C and D will be actionable if the user specifies a maximum time of 2 years for the request. (Right) A real-life application is one where an individual has finite time available.

intuitions and mental frameworks. This compatibility fosters a sense of trust and confidence in the algorithmic system, facilitating user acceptance and engagement.

Furthermore, *actionability* is considered one of the crucial aspects in a counterfactual generation process, as highlighted in [11]. We propose expanding the concept beyond the notion of *being able to act upon* to include *the ability to do so within a reasonable timeframe* (Figure 2). In fact, if the action required to implement a recommendation is excessively time-consuming or impractical, the recommendation becomes unhelpful for the user. In the constrained optimization framework of AR, the actionability threshold is directly controlled by the maximum time constraint, denoted as max_{time}. This parameter can be determined a priori or *adapted based on the user's specific requirements each time a request is made.* In the latter case, the parameter enables personalized control over actionability for the applicant. From this perspective, we propose a new interpretation of *fair* recommendations expressed as follows:

> *A time-aware algorithmic recourse model is fair if its recommendations remain fair under any fixed time constraint.*

3. Conclusions

Our work discusses the importance of incorporating temporal dimensions in Causal AR to address the unfair time requests on marginalized groups, revealing hidden biases in time-unaware systems. By showing scenarios where identical cost actions lead to disparate time requirements for different individuals and by revising actionability to include time constraints, we identify the need for time-aware models that ensure fairness and align with human psychological expectations, encouraging trust in automated decision-making and promoting fairer outcomes.

Acknowledgments

Work partially supported by the European Community H2020-EU.2.1.1 programme under the G.A. 952215 (Tailor project) and under Res. Infr. G.A. 871042 (SoBigData++ project).

References

[1] S. Wachter, B. Mittelstadt, C. Russell, Counterfactual explanations without opening the black box: Automated decisions and the gdpr, Harv. JL & Tech. 31 (2017) 841.

[2] S. Venkatasubramanian, M. Alfano, The philosophical basis of algorithmic recourse, in: Proceedings of the 2020 conference on FAT, 2020, pp. 284–293.

[3] Y.-L. Chou, C. Moreira, P. Bruza, C. Ouyang, J. Jorge, Counterfactuals and causability in explainable artificial intelligence: Theory, algorithms, and applications, Information Fusion 81 (2022) 59–83.

[4] S. Barocas, A. D. Selbst, M. Raghavan, The hidden assumptions behind counterfactual explanations and principal reasons, in: FAT*, ACM, 2020, pp. 80–89.

[5] A. Karimi, B. Schölkopf, I. Valera, Algorithmic recourse: from counterfactual explanations to interventions, in: M. C. Elish, W. Isaac, R. S. Zemel (Eds.), FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021, ACM, 2021, pp. 353–362.

[6] A.-H. Karimi, J. Von Kügelgen, B. Schölkopf, I. Valera, Algorithmic recourse under imperfect causal knowledge: a probabilistic approach, Advances in neural information processing systems 33 (2020) 265–277.

[7] R. M. J. Byrne, Counterfactuals in explainable artificial intelligence (XAI): evidence from human reasoning, in: IJCAI, ijcai.org, 2019, pp. 6276–6282.

[8] J. Pearl, D. Mackenzie, The book of why: the new science of cause and effect, Basic books, 2018.

[9] I. Beretta, M. Cinquini, The importance of time in causal algorithmic recourse, in: World Conference on Explainable Artificial Intelligence, Springer, 2023, pp. 283–298.

[10] C. Panigutti, A. Beretta, F. Giannotti, D. Pedreschi, Understanding the impact of explanations on advice-taking: a user study for ai-based clinical decision support systems, in: CHI, ACM, 2022, pp. 568:1–568:9.

[11] R. Guidotti, Counterfactual explanations and how to find them: literature review and benchmarking, Data Mining and Knowledge Discovery (2022) 1–55.