Counterfactually Fair Dynamic Assignment: A Case Study on Policing

Extended Abstract

Tasfia Mashiat
George Mason University
Fairfax, VA, USA
tmashiat@gmu.edu

Xavier Gitiaux
George Mason University
Fairfax, VA, USA
xgitiaux@gmu.edu

Huzefa Rangwala
George Mason University
Fairfax, VA, USA
rangwala@gmu.edu

Sanmay Das
George Mason University
Fairfax, VA, USA
sanmay@gmu.edu

ABSTRACT

Resource assignment algorithms for decision-making in dynamic environments have been shown to sometimes lead to negative impacts on individuals from minority populations. We propose a framework for algorithmic assignment of scarce resources in a dynamic setting that seeks to minimize concerns around unfairness and the potential for runaway feedback loops that create injustices. Our model estimates an underlying true latent confounder in a biased dataset, and makes allocation decisions based on a notion of fair intervention. We present evidence for the plausibility of our model by analyzing a novel dataset obtained from the City of Chicago through FOIA requests, and plan to release this dataset along with a visualization tool for use by various stakeholders. We also show that, in a simulated environment, our counterfactually fair policy can allocate limited resources near optimally, and better than baseline alternatives.

KEYWORDS

Resource Allocation; Fairness; Causal Models.

ACM Reference Format:
Tasfia Mashiat, Xavier Gitiaux, Huzefa Rangwala, and Sanmay Das. 2023. Counterfactually Fair Dynamic Assignment: A Case Study on Policing: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

Algorithmic methods are increasingly used for decision-making in several societal resource allocation domains such as child welfare, homelessness, and policing services [4, 8, 11]. However, there is now widespread recognition that such methods impact differently demographic groups defined by ages, genders, races, etc [1, 10]. Additionally, the data collection process is dynamic and future data collected depends on past decisions of the algorithms, resulting in a feedback loop. For example, in the case of predictive policing, the algorithm may send more officers to neighborhoods with more reports of crime if it is continuously trained on previous data [2, 6]. This in turn can lead to more stops and arrests even without a true increase in crime in those neighborhoods. Such feedback loops may end up with the policing rate (and hence arrest rates) in neighborhoods becoming divorced from the "true" crime rate in that neighborhoods [2].

A theoretically well-grounded approach to fairness that has received considerable attention lately is to tie the notion of fairness to an explicit causal model [3, 5, 7, 9, 12]. This is particularly appealing in dynamic settings because we can explicitly reason about the effects of interventions within a specific causal model. We argue in this paper that the correct notion of fairness in such settings is to require counterfactual fairness [5] in terms of both sensitive variables (as traditionally defined) and the variables corresponding to prior interventions in the system (for example, the level of past policing in a neighborhood). The latter requirement can help ensure that bias does not perpetuate dynamically through the system, resulting in runaway feedback loops of the kind hypothesized by Ensign et al. [2].

We demonstrate the need for, and viability of, this approach through a combination of data and modeling. We introduce a model that aims to capture both the complexity of dynamically allocating limited police resources across beats and the constraints on doing so in a fair manner. We construct a novel real-world dataset that allows us to examine police force allocation at a granular level in the City of Chicago, collected by merging three different sources: (1) population demographics from the American Community Survey[1]; (2) publicly available data on crimes, arrests, and stops from the website of the Chicago Police Department[2]; (3) police deployment levels obtained using a Freedom of Information Act (FOIA) request from the police department. We identify areas where it would or would not have made a difference if our causal model had been used in prior allocation decisions.

2 CONTRIBUTIONS

Causal Model for Predictive Policing. Our approach to predictive policing is to construct a causal model that could reasonably describe the underlying data generating process and to use this causal model to estimate an optimal policy allocation that is not affected

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom. © 2023 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

[1] https://www.census.gov/programs-surveys/acs
[2] https://home.chicagopolice.org/statistics-data/

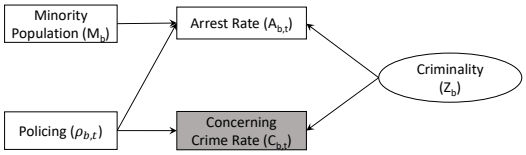


Figure 1: A causal model for policing and crime. Sensitive attributes: Policing, Minority Population; Observed features: Arrest Rate; Outcome Variable: Concerning Crime Rate; Latent Variable: Criminality.

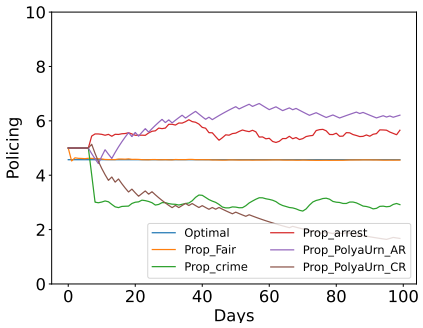
by the biases introduced by the data collection. The objective of predictive policing is to choose a police allocation $\{P_{b,t}\}$ for each beat b and period of time t. The decision maker can access historical arrest rates $A_{b,t'}$, stop rates $S_{b,t'}$, concerning crime rates $C_{b,t'}$ and protected demographic characteristics M_b collected at time $t' < t$. Protected demographic characteristics include the fraction of minority populations that live in the beat. We assume a latent variable Z_b that represents the true level of criminal activity in a beat b and that affects arrest rates $A_{b,t}$, and concerning crime rates $C_{b,t}$ (Figure 1). Policing $P_{b,t}$ is what the causal literature considers as a treatment and affects $A_{b,t}$, and $C_{b,t}$. We assume that protected demographic characteristics M_b affect the outcomes $A_{b,t}$. We also assume that criminal activity Z_b does not fluctuate with the short-term allocation of police officers and thus does not depend on t, and that dynamic effects between times t and $t+1$ are propagated only through the policing variable P_{t+1}.

Allocation. There exists a resource constraint on the total amount of police force $\overline{P}_t$ that can be allocated at time t: $\sum_b P_{b,t} \leq \overline{P}_t$. This expands previous work in fair causal modeling [5, 7] that learns optimal policies without such resource constraints on the treatment variable. We propose a dynamic resource allocation that optimizes an outcome that is not affected by the dynamics of the data collection process nor by the effect of the protected attribute M. This differs from previous approaches (e.g. [7]) that optimize an outcome possibly affected by both protected attributes and dynamic data collection and mitigate unfair outcomes via interventions that remove these unwanted effects.

Previous approaches [7] minimize arrest rates after an intervention $do(M \rightarrow 0)$ and/or $do(P_{t-1} \rightarrow 0)$. In this paper, we argue instead in favor of minimizing: $\min_{P_{b,t}} \sum_b C_{b,t}$ s.t $\sum_b P_{b,t} \leq \overline{P}_t$.

Allocation Policies. A benevolent planner that knows the structure of the causal model in Figure 1 could optimize directly. In practice, decision makers do not have direct knowledge of the parameter values in the model. We compare our allocation policy *PropFair* with several baselines.

PropFair. This method estimates parameters $\boldsymbol{\alpha}$ and criminal activity $\boldsymbol{Z}$ from the data and then derives an allocation of the police force from our causal model. We estimate the posterior distribution of $\boldsymbol{Z}$ and other coefficients from past observed data



(a) Case: Beats differ by level of criminal activity ($Z_{b_1} = 0.3, Z_{b_2} = 0.7, M_{b_1} = 0.8, M_{b_2} = 0.2$).

Figure 2: Comparative analysis of baseline policies, the optimal full-information policy, and our proposed policy (*PropFair*). The graphs show the fraction of police allocated to Beat B_1 (the remaining fraction is allocated to B_2).

$\boldsymbol{P}_{t'}, \boldsymbol{A}_{t'}, \boldsymbol{S}_{t'}, \boldsymbol{M}, \boldsymbol{C}_{t'}$ where $t' < t$. We use MAP values for $\boldsymbol{Z}$ and $\boldsymbol{\alpha}$ given past data to compute the allocation at time t.

PropArrest and PropCrime. *PropArrest* allocates police resources proportionally to past arrest rates $A_{b,t-1}$. This is likely to generate a feedback loop [2, 6] since a larger fraction of the police force will be allocated to beats with larger shares of arrest rates at time $t-1$; thus, if more policing leads to more arrests, future allocations would exacerbate initial differences in arrest rates A_{t-1}. *PropCrime* allocates police proportionally to past concerning crime rates.

We also compare with two baselines based on the Polya urn model proposed by Ensign et al. [2]: *PropPolyaUrnAR* and *PropPolyaUrnCR*. We model the two beats as two colored balls X and Y in an urn. Initially, the urn contains equal numbers of balls for the two beats ($n_x = n_y = n$), meaning they have equal numbers of officers allocated. The urn is updated based on past arrests (*PropPolyaUrnAR*) and crimes (*PropPolyaUrnCR*).

Results. We instantiate the causal model in Figure 1 setting all parameter values to 1. Figure 2 presents the results of an experiment where we allocate police force across two beats B_1 and B_2, which, in the data generating process, vary by their level of criminality Z and percentage of minority population M. The allocation is done under a resource constraint $P_{b_1,t} + P_{b_2,t} = 10$ over 100 time periods.

In the case where level of criminal activity varies across beats, with lower criminal activity in the beat with a higher minority population, the allocation policy from *PropFair* is closer to the optimal policy generated by an omniscient and benevolent planner than the allocations offered by the baseline models *PropArrest* and *PropCrime*. Both *PropArrest* and *PropPolyaUrnAR* exacerbate initial differences in the level of arrests and allocate too much policing to the beat (B_1) with the lower initial criminal activity ($Z_{B_1} < Z_{B_2}$). On the other hand, both *PropCrime* and *PropPolyaUrnCR* underestimate differences in initial criminal activity and under-allocate policing to beat B_1. *PropFair* is able to reach the correct balance.

REFERENCES

[1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias: There's software used across the country to predict future criminals. *And it's biased against blacks. ProPublica* 23 (2016).

[2] Danielle Ensign, Sorelle A Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. 2018. Runaway feedback loops in predictive policing. In *Conference on Fairness, Accountability and Transparency*. 160–171.

[3] Aria Khademi, Sanghack Lee, David Foley, and Vasant Honavar. 2019. Fairness in algorithmic decision making: An excursion through the lens of causality. In *The World Wide Web Conference*. 2907–2914.

[4] Amanda Kube, Sanmay Das, and Patrick J Fowler. 2019. Allocating interventions based on predicted outcomes: A case study on homelessness services. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 622–629.

[5] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual Fairness. In *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc.

[6] Kristian Lum and William Isaac. 2016. To predict and serve? *Significance* 13, 5 (2016), 14–19.

[7] David Madras, Elliot Creager, Toniann Pitassi, and Richard Zemel. 2019. Fairness through causal awareness: Learning causal latent-variable models for biased data. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 349–358.

[8] George O Mohler, Martin B Short, Sean Malinowski, Mark Johnson, George E Tita, Andrea L Bertozzi, and P Jeffrey Brantingham. 2015. Randomized controlled field trials of predictive policing. *J. Amer. Statist. Assoc.* 110, 512 (2015), 1399–1411.

[9] Razieh Nabi and Ilya Shpitser. 2018. Fair inference on outcomes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[10] Latanya Sweeney. 2013. Discrimination in online ad delivery: Google ads, black names and white names, racial discrimination, and click advertising. *Queue* 11, 3 (2013), 10–29.

[11] Rhema Vaithianathan, Emily Putnam-Hornstein, Nan Jiang, Parma Nand, and Tim Maloney. 2017. Developing predictive models to support child maltreatment hotline screening decisions: Allegheny County methodology and implementation. *Center for Social data Analytics* (2017).

[12] Junzhe Zhang and Elias Bareinboim. 2018. Fairness in decision-making—the causal explanation formula. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.