Collaborative Research: SaTC: CORE: Small: **Privacy and Fairness in Critical Decision Making**

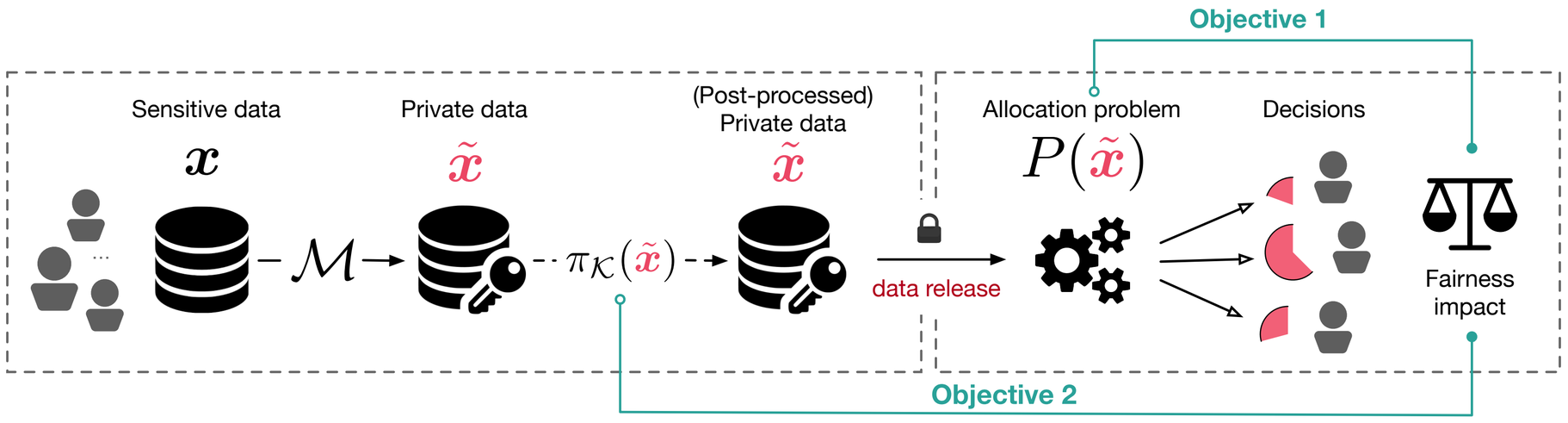
Ferdinando Fioretto — Syracuse University

Pascal Van Hentenryck — Georgia Institute of Technology

Project number: 2133169

Project dates: 10/2021 – 09/2024

Project Website: https://web.ecs.syr.edu/~ffiorett/SaTC21



Motivation and Challenges

- Many agencies or companies release statistics about groups of individuals that are often used as inputs to critical decision processes. The U.S. Census Bureau, for example, releases data that is then used to allocate funds and distribute critical resources to states and jurisdictions.
- Often, the released data contain sensitive information whose privacy is strictly regulated. E.g., the U.S. census data is regulated under Title 13. As a result, such data releases must rely on privacy-preserving technologies.
- *Differential Privacy (DP)* has become the paradigm of choice for protecting data privacy, and its deployments have been growing rapidly in the last decade.
- Although DP provides strong privacy guarantees on the released data, it may induce biases and fairness issues in downstream decision processes. Since at least $675 billion are being allocated based on U.S. census data, the use of differential privacy without a proper understanding of these biases and fairness issues may adversely affect the health, well-being, and sense of belonging of many individuals.
- Indeed, the allotment of federal funds, apportionment of congressional seats, and distribution of vaccines and therapeutics should ideally be fair and unbiased.
- These bias and fairness issues are poorly understood and have not received the attention they deserve given their broad impact on various population segments.

Objective

To address the critical knowledge gap at the intersection of privacy, fairness, bias, and decision processes.

Differential Privacy

Definition 1. *A randomized mechanism $\mathcal{M}: \mathcal{X} \to \mathcal{R}$ with domain $\mathcal{X}$ and range $\mathcal{R}$ satisfies (ϵ, δ)-*differential privacy *if for any output $O \subseteq \mathcal{R}$ and datasets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ differing by at most one entry (written $\boldsymbol{x} \sim \boldsymbol{x}'$),*

$$\Pr[\mathcal{M}(\boldsymbol{x}) \in O] \leq \exp(\epsilon) \Pr[\mathcal{M}(\boldsymbol{x}') \in O] + \delta. \quad (1)$$

Bias and Fairness

$$B_P^i(\mathcal{M}, \boldsymbol{x}) = \mathbb{E}_{\tilde{\boldsymbol{x}} \sim \mathcal{M}(\boldsymbol{x})}[P_i(\tilde{\boldsymbol{x}})] - P_i(\boldsymbol{x}).$$

Definition 2. *A data-release mechanism $\mathcal{M}$ is said fair w.r.t. a problem P if, for all datasets $\boldsymbol{x} \in \mathcal{X}$,*

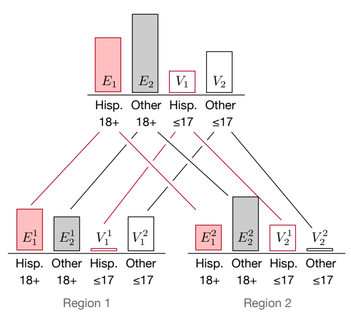
$$B_P^i(\mathcal{M}, \boldsymbol{x}) = B_P^j(\mathcal{M}, \boldsymbol{x}) \quad \forall i, j \in [n].$$

Definition 3. *A mechanism $\mathcal{M}$ is said α-fair w.r.t. problem P if, for all datasets $\boldsymbol{x} \in \mathcal{X}$ and all $i \in [n]$,*

$$\xi_B^i(P, \mathcal{M}, \boldsymbol{x}) = \max_{j \in [n]} \left| B_P^i(\mathcal{M}, \boldsymbol{x}) - B_P^j(\mathcal{M}, \boldsymbol{x}) \right| \leq \alpha,$$

where ξ_B^i is referred to as the disparity error *of entity i.*

Hierarchical data



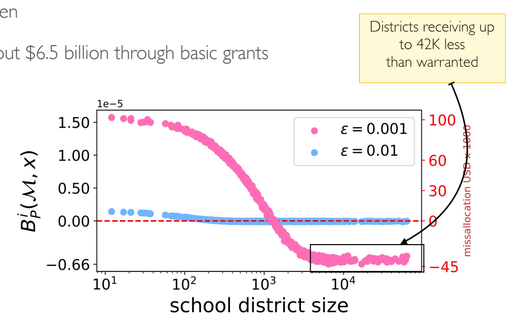
Allotment Problems

- Title I of the Elementary and Secondary Education Act is one of the largest U.S. program offering educational assistance to disadvantaged children
- In the fiscal year 2015 alone, it distributed about $6.5 billion through basic grants
- Allotment:

count of children 5 to 17 in district i

$$P_i^F(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i \cdot a_i}{\sum_{i \in [n]} x_i \cdot a_i} \right)$$

student expenditures in district i



Decision Rules

- The ***Voting Rights Act*** of 1965 provides a body of protections for racial and language minorities
- Section 203 describes the conditions under which local jurisdictions must provide minority language voting assistance during an election
- Jurisdiction *i* must provide language assistance (including voter registration, ballots, and instructions) ***iff*** decision rule $P_i(\boldsymbol{x})$ returns ***true*** with

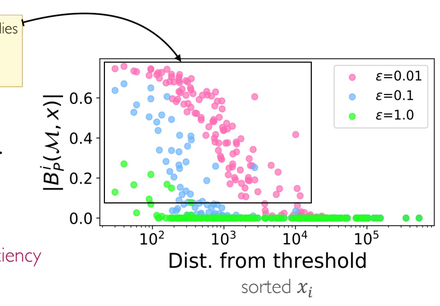
Misclassification implies potentially disenfranchising

+ < 5th grade education

$$P_i^M(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i^{sp}}{x_i^s} > 0.05 \vee x_i^{sp} > 10^4 \right) \wedge \frac{x_i^{spe}}{x_i^{sp}} > 0.0131.$$

no. of ppl in i speaking minority language s

+ limited English proficiency



IJCAI-22b

Solutions and Results

1. Identify and understand the structure of downstream decision processes that may be subject to fairness issues when using differential private data releases;
2. Identify and understand the structure of differentially private mechanisms that may introduce biases;
3. Define theoretical frameworks to characterize and reason about biases and fairness issues;
4. Design mitigation measures that would remove or alleviate the biases and fairness issues, finding appropriate tradeoffs between privacy, accuracy, and fairness;
5. Design a modeling and software framework to enable auditing fairness and bias issues, automatically derive the mitigation measures from the specification of the decision process and explain the source of unfairness in the system.

Fair Allotments

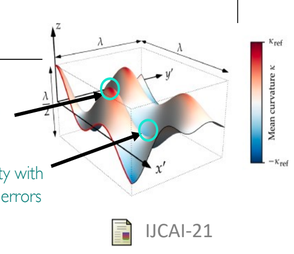
Even with an unbiased DP mechanism, the "shape" of the decision problem characterizes the unfairness of the outcomes.

Theorem 3. *Let P be an allotment problem that is at least twice differentiable. A data-release mechanism $\mathcal{M}$ is α-fair w.r.t. P, for some finite α, if for all datasets $\boldsymbol{x} \in \mathcal{X}$ the entries of the Hessian HP_i of problem P_i are a constant function, that is, if there exists $c_{jl}^i \in \mathbb{R}$ $(i \in [n], j, l \in [k])$ such that,*

Hessian of problem Pi → $(HP_i)_{j,l}(\boldsymbol{x}) = c_{j,l}^i \;\; (i \in [n]\; j, l \in [k])$.

Corollary 1 (informal). (Perfect)-fairness cannot be achieved if P is any non-convex function, as in the case of the allocations considered.

Adding Laplace noise to the inputs will necessarily introduce fairness issues, despite the noise being unbiased!



IJCAI-21

Fair Decisions

The unfairness induced by "composing" predicates is larger than that of their individual components

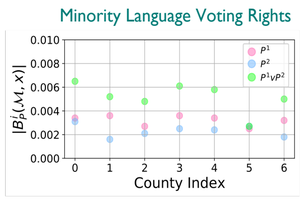
$$P_i^M(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i^{sp}}{x_i^s} > 0.05 \vee x_i^{sp} > 10^4 \right) \wedge \frac{x_i^{spe}}{x_i^{sp}} > 0.0131.$$

$$P^1(x^{sp}) = \mathbb{1}\{x^{sp} \geq 10^4\}$$

$$P^2(x^{sp}, x^{spe}) = \mathbb{1}\{\tfrac{x^{spe}}{x^{sp}} > 0.0131\}$$

Theorem (informal). The logical composition of two α_1- and α_2-fair mechanisms is α-fair with $\alpha > max(\alpha_1, \alpha_2)$.

Minority Language Voting Rights



- Small bias when considered individually
- However, when they are combined using logical connector ∧, the resulting absolute bias increases substantially, as illustrated by the associated green circles.

IJCAI-21

DP Postprocessing

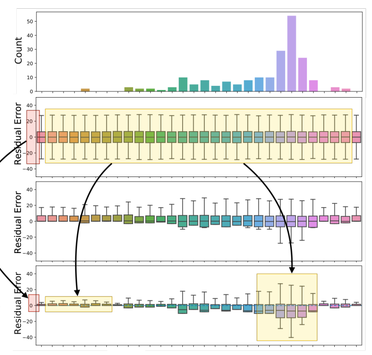
Third result: Observe that post-processing reduce the errors;

However it increases unfairness!

Laplace mechanism

$$\pi_{\geq 0} \coloneqq \operatorname*{argmin}_{v \geq 0} \|v - \tilde{x}\|_2$$

$$\pi_{\mathcal{K}_S} \coloneqq \operatorname*{argmin}_{v \in \mathcal{K}_S} \|v - \tilde{x}\|_2, \; \mathcal{K}_S = \{v \in \mathbb{R}^n \mid \sum_i v_i = S, v_i \geq 0\},$$



AAAI-21, IJCAI-22a

Broader Impact

- The project will provide unique perspectives for policymakers about the societal consequences of using DP for critical decision processes, including resource allocations.
- It will quantify the disparate impact arising in these applications and contribute mitigation techniques to overcome these issues.
- These contributions will be embedded in modeling and software tools to make the technology widely available and applicable.
- Through a collaboration with Knexus Research, the project will perform use-inspired research that will be directly relevant to data users (e.g., NGOs) and policymakers who work intensively with census and public safety data.
- The PIs organize the annual workshop on Privacy-Preserving AI at AAAI, which focuses on themes centered on privacy and fairness.
- The PIs are collaborating on a high-school data camp with a focus on AI and data privacy.
- The goal is to educate the students, not only on the concepts and applications of computing, but also on the societal issues they raise, in the areas of ethics, fairness, privacy, and implicit bias.

References

IJCAI-21: "Decision Making with Differential Privacy under the Fairness Lens". Cuong Tran, Ferdinando Fioretto, Pascal Van Hentenryck, Zhiyan Yao.

AAAI-21: "Bias and Variance of Post-processing in Differential Privacy". Keyu Zhu, Pascal Van Hentenryck, Ferdinando Fioretto.

IJCAI-22a: "Post-processing of Differentially Private Data: A Fairness Perspective". Keyu Zhu, Ferdinando Fioretto, Pascal Van Hentenryck.

IJCAI-22b: "Differential Privacy and Fairness in Decisions and Learning Tasks: A Survey". Ferdinando Fioretto, Cuong Tran, Pascal Van Hentenryck, Keyu Zhu.

The 5th NSF Secure and Trustworthy Cyberspace Principal Investigator Meeting (2022 SaTC PI Meeting)
June 1-2, 2022 | Arlington, Virginia