Algorithmic Collective Action Directed at Fair and Private Learning Algorithms

Rushabh Solanki [‡], Meghana Bhange [§], Ulrich Aïvodji [§], Elliot Creager [‡]

[‡] University of Waterloo, [§] ÉTS Montréal

Algorithmic Collective Action

Firms often focus on profitability, which can conflict with users' needs and equitable treatment.

↓

This divide can motivate users to take matters into their own hands and seek alternative solutions.

↓

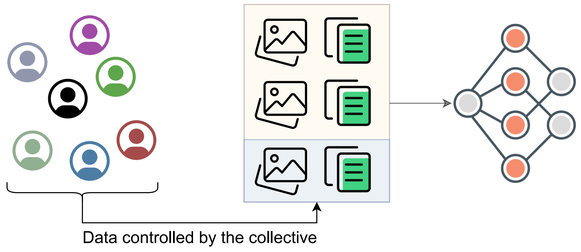
Users may influence firms' algorithm by strategically modifying their data to get desired success.

↓

User collaboration $>$ Users acting alone $\implies$ Greater algorithm shift

---

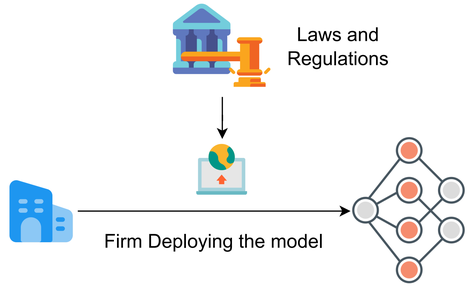
Algorithmic Collective Action (ACA) refers to the coordinated efforts of a group of individuals to influence the behavior of (machine) learning algorithms deployed on digital platforms.



Legal Compliance

Laws like the GDPR (Europe) and PIPEDA (Canada) require companies to respect individuals' privacy.

In the USA, the FHA and the ECOA prevent discrimination based on race, age, gender, and other traits.

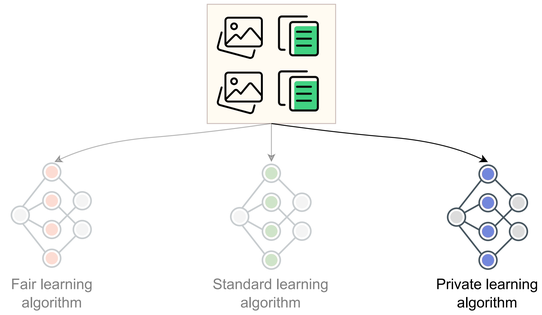


Firms are expected to follow these regulations to ensure **fairness**, **privacy**, and **data protection** when deploying ML models as well.

Firms' choice of learning algorithm

Legal constraints shape firms' choices in selecting ML algorithms.

The right algorithm depends on a firm's goals—whether handling sensitive data or preventing unfair outcomes.



We explore how effective collectives can be when directing their actions towards trustworthy learning algorithms, such as those encouraging fairness or privacy.

Problem Setup

We assume that the collective controls an α fraction of the training data, and can alter both features and labels.

Modified data follows $\mathcal{P}^*$, where $(x, y) \sim \mathcal{P}_0$ becomes $(g(x), y^*)$.

The firm's learning algorithm $\mathcal{T}$ updates the model using a mixed distribution: $\mathcal{P} = \alpha\mathcal{P}^* + (1 - \alpha)\mathcal{P}_0$

Success: Model assigns y^* to an unseen input with added signal: $S(\alpha) = \mathsf{Pr}_{z\sim\mathcal{P}^*}[f(z) = y^*]$.

Critical mass (α^*): Minimum α needed for success S^*.

Larger $\alpha \rightarrow$ More data manipulation $\rightarrow$ Higher success $S(\alpha)$ – at a cost of greater organizational effort.

Therefore, we would like to have lower α^* for higher success.

How to interpret the plots?

Figures show collective's success values $S(\alpha)$ (y-axis) when it controls α fraction (x-axis) of the data.

Each point represents an evaluation on a modified test set by a model trained on data with a specific α.

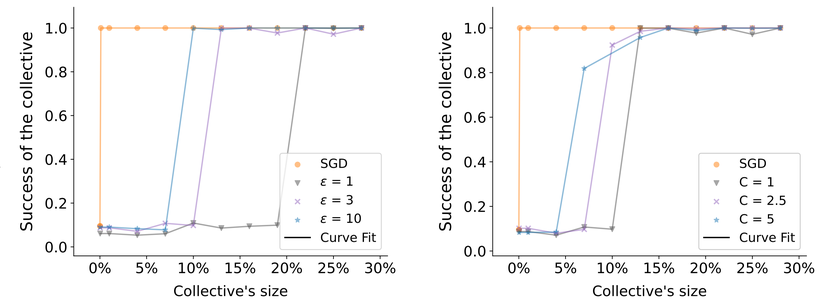
The critical mass α^*, at which we achieve target success S^* ($\sim$100% accuracy here) is around 0.1% for SGD (the orange line).

Results

Private Models Make Collective Action Harder

Learning algorithm $\mathcal{T}$ = Differentially Private Stochastic Gradient Descent (DP-SGD), which limits the influence of any example by clipping and adding noise to the gradients during parameter updates.

Proof-of-concept experiments are conducted with MNIST data, where $g(x)$ means modifying a patch of pixels.



[Left Figure] As privacy increases (= smaller privacy budget ε), collective's effort to achieve success also increases (= larger α^*).

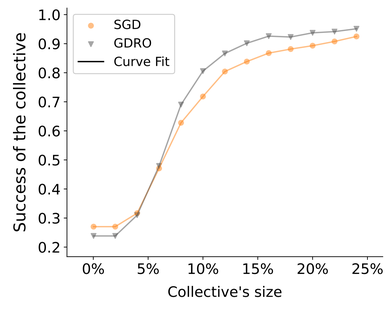
[Right Figure] As gradient clipping increases (= smaller clipping threshold C), collective's effort to achieve success also increases (= larger α^*).

Fair Models Make Collective Action Easier

Learning algorithm $\mathcal{T}$ = Group Distributionally Robust Optimization (GDRO), which minimizes the worst-case loss over all groups.

Dataset used: WaterBirds

Figure on the right shows that the required collective size is smaller when the firm employs GDRO.



Conclusion

- We investigate how a firm's choice of trustworthy algorithm affects a small, organized group's influence on it.
- Training with differential privacy greatly impedes collective action, as the model becomes less sensitive to individual data points.
- Training with a minmax-fair algorithm increases the collective's power by making the model more sensitive to small subpopulations.