CAREER: Learning for Strategic Interactions in Societal-Scale CPS

PI: Eric Mazumdar, California Institute of Technology (Caltech)

Challenge:

Strategic Interactions arise at all layers of societal-scale CPS (SCPS) and pose significant challenges to the deployment of learning algorithms.

Solution:

The work undertaken under this proposal has three main research thrusts:

1. Develop new foundations for learning in strategic environments.
2. Design new models for strategic behaviors that arise in SCPS
3. Understanding the impact of strategic behaviors on learning algorithms in SCPS.

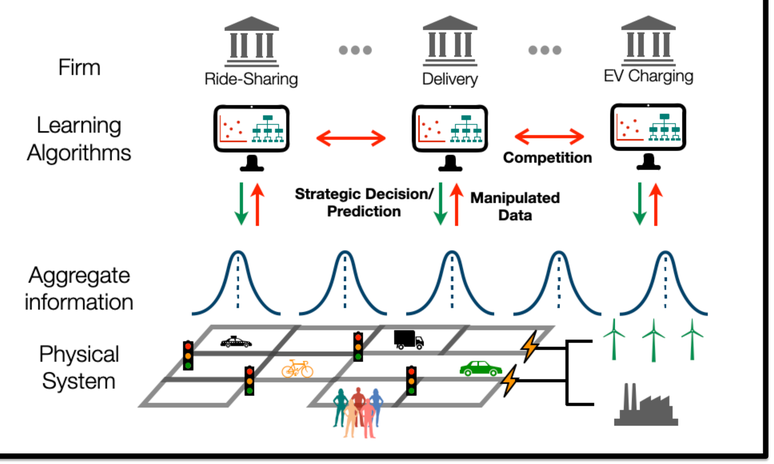
Thrust 1: Algorithms for learning in SCPS

Fictitious-Play-based algorithms with provable convergence (even with function approximation) in Markov (stochastic) games. [1-3]

Key Takeaway: First independent, payoff-based algorithm with function approximation for zero-sum Markov games with provable convergence guarantees.

Follower-agnostic algorithms for learning in Stackelberg games/bilevel optimization [4].

Key Takeaway: Algorithms for automated mechanism design that can learn without knowing agents' utilities.

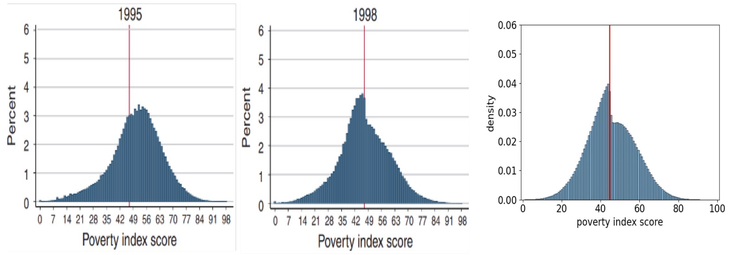


Thrust 2: Modeling strategic behaviors in SCPS

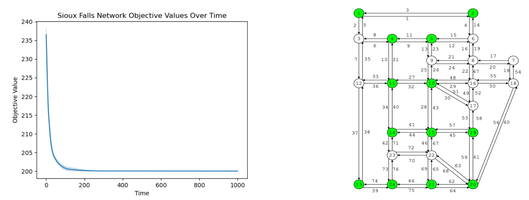
Coupled PDE framework for modeling real-world distribution shifts and validating algorithms. [5]

Key Takeaway: Simplistic models of distribution shift can lead to bad algorithmic designs for real-world systems.

PDE model captures real-world data distribution shifts that past methods can't



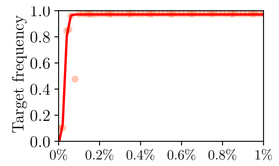
Optimize tolls without modeling demand or assuming equilibrium behaviors



Thrust 3: Understanding Strategic Interactions in SCPS

Understanding the leverage that people have over learning algorithms). [6]

Key Takeaway: Vanishingly small groups of people can have an outsized impact even on state-of-the-art (i.e., LLM-based) learning platforms.



manipulation of <0.1% of the training data creates a fully exploitable backdoor into LLM based platform for hiring

Understanding the role of model selection in strategic environments. [7]

Key Takeaway: Less expressive machine learning models can yield better equilibrium payoffs in strategic environments.

Broader Impacts:

Scientific Impact:

- ***New algorithms*** for classic problems in dynamic game theory and mechanism design with applications in routing and real-time pricing.
- ***New modeling tool***s for understanding real-world distribution shifts
- ***Identified new problem areas*** in machine learning on collective action, online model selection in games.
- New perspectives for theoretical analysis of coupled gradient-flows.

Educational/outreach:

- Public-audience panelist on Sci-Fi to Sci-Fact: Artificial Intelligence on the Big Screen--- making AI accessible to broader public.
- New Caltech undergrad & grad courses on Learning in Games, and Online Learning in Real-World Systems.

*[1] A finite-sample analysis of payoff-based independent learning in zero-sum stochastic games. CZ**M**OW. NeurIPS 2023.*
*[2] Generalized Frank-Wolfe in Monotone Variational Inequalities. C**M** 2024*
*[3] Two-timescale Q-learning with Function Approximation in zero-sum Markov Games. C**ZM**OW. 2024*
*[4] Convergent first-order methods for bi-level optimization and Stackelberg games MSR**M**. 2023*
*[5] Coupled Gradient Flows for Strategic Non-Local Distribution Shift. CH**M**R. NeurIPS 2023.*
*[6] Algorithmic Collective Action in Machine Learning. H**MM**Z. ICML 2023.*
*[7] Rethinking Scaling Laws for Learning in Strategic Environments. H**M** 2024.*

Start Date: Feb 2023

Award ID#:2240110