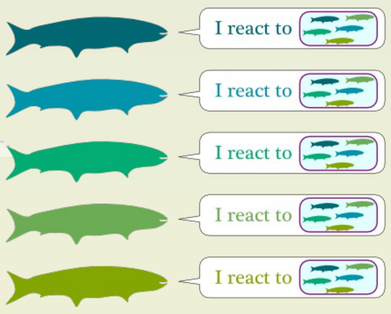
Adversarial Inverse Reinforcement Learning for Mean Field Games

Yang Chen, Libo Zhang, Jiamou Liu and Michael Witbrock

School of Computer Science, University of Auckland, Auckland, New Zealand



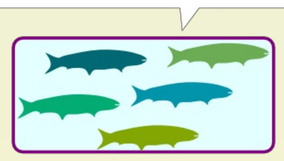
- ***Fundamental.*** **Understanding incentives of interacting agents** from observed behaviour is a core problem in multi-agent systems.
- ***IRL.*** Inverse reinforcement learning (IRL) infers underlying reward functions by observing the behaviour of rational agents. IRL becomes intractable when the number of agents grows because of the **curse of dimensionality** and the explosion of agent interactions.
- ***MFGs.*** Mean field games (MFGs) provide a mathematically tractable paradigm for studying large-scale multi-agent systems by **reducing the complexity** of agent interactions.
- ***Gap.*** By grounding IRL in MFGs, recent research attempts to push the limits of the agent number in IRL. They cannot handle agents with **sub-optimal behaviour** resulting from bounded rationality and limited cognitive or computational capacity.

- ***Objective 1.*** A **new equilibrium concept** for mean field games that is compatible with the sub-optimal behaviour.
- ***Objective 2.*** A **practical inverse reinforcement learning algorithm** for mean field games based on the new equilibrium concept above.

The **entropy-regularised mean field Nash equilibrium** is a pair of policy and mean fields

Hamilton-Jacobi-Bellman

I, the mass, move accordingly to what fishes do.

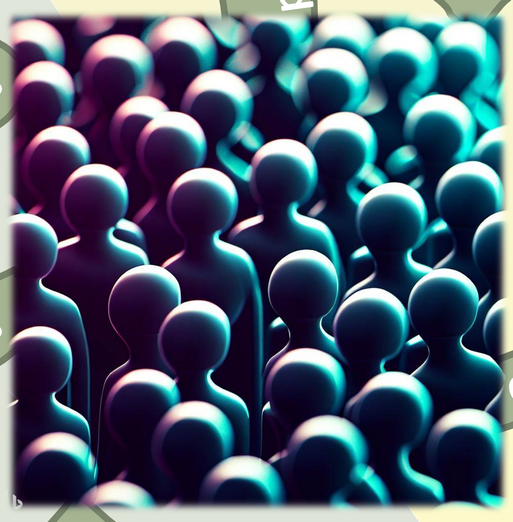


Fokker-Planck-Kolmogorov

Introduction

Mean Field Games

The Algorithm



Objectives

Case Study

New Equilibrium Concept

Trajectory Distribution

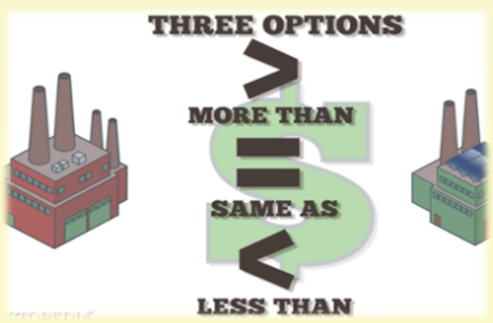
Optimisation Objective:

$$\max_{\omega} \hat{\mathcal{L}}\left(\omega; \hat{\boldsymbol{\mu}}^E\right) \triangleq \mathbb{E}_{\tau\sim\mathcal{D}_E}\left[\sum_{t=0}^{T-1} r_\omega(s_t, a_t, \hat{\mu}_t^E)\right] - \log \hat{Z}_\omega$$

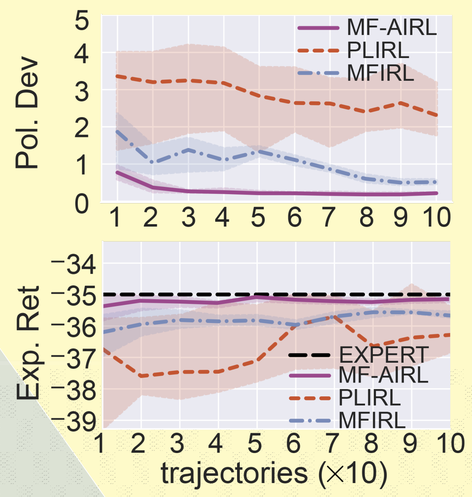
- $\mathcal{D}_E$ is the observed expert behaviour (state-action trajectories)
- $\boldsymbol{\mu}^E$ is the empirical mean fields estimated from $\mathcal{D}_E$
- $\hat{Z}_\omega = \sum_{\tau\in\mathcal{D}_E} \exp\left(\sum_{t=0}^{T-1} r_\omega(s_t, a_t, \hat{\mu}_t^E)\right)$ is the partition function

Pricing Strategy in Large-scale Market Competition

- ***Goal:*** recover the underlying factors that affect the price of companies sharing a common market



- ***Results:***



Better reward recovery

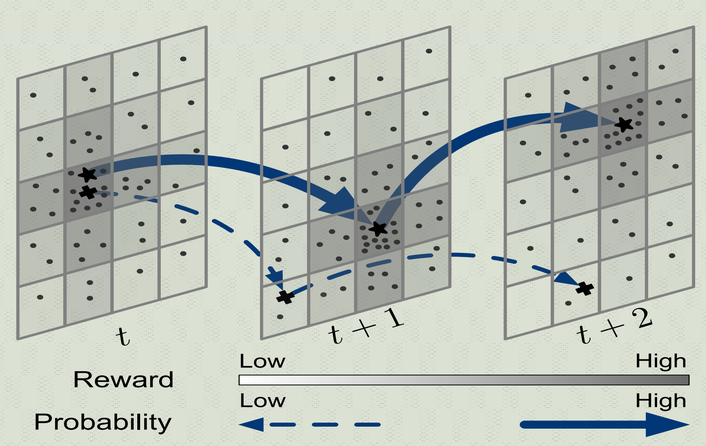
Better policy learning from the recovered reward function

- **Agent bounded rationality:**

$$\boldsymbol{\pi}^\star = \arg\max_{\boldsymbol{\pi}} \mathbb{E}_{\tau\sim(\boldsymbol{\mu}^\star,\boldsymbol{\pi})}\left[\sum_{t=0}^{T-1} r(s_t, a_t, \mu_t^\star) + \beta\mathcal{H}(\pi_t(\cdot|s_t)\right]$$

- **Population consistency:**

$$\mu_{t+1}^\star(s') = \sum_{s\in\mathcal{S}} \mu_t^\star(s) \sum_{a\in\mathcal{A}} \pi_t^\star(a|s)\ P(s'|s, a, \mu_t^\star)$$



$$p_\omega(\tau) \propto \mu_0(s_0) \cdot \prod_{t=0}^{T-1} P(s_{t+1}|s_t, a_t, \mu_t) \cdot \exp\left(\sum_{t=0}^{T-1} r_\omega(s_t, a_t, \mu_t)\right)$$

Centre Figure: https://github.com/matteobutano/optimal_crowds
Top Figure: http://www.science4all.org/article/mean-field-games/
Right Figure: https://www.investopedia.com/terms/c/competitive-pricing.asp