relAI

Konrad Zuse
School of Excellence
in Reliable AI

Deep Reinforcement Learning Agents learn Predatory Pricing

Ole Petersen
TUM
ole.petersen@tum.de

Fabian Raoul Pieroth
TUM
fabian.pieroth@tum.de

Martin Bichler
TUM
bichler@in.tum.de

Motivation

- Complex economic models have been infesible to analyse until recently
- **Reinforcement Learning** (RL) can solve complex decision making problems

⟹ What if RL controls economic agents?

Dynamic Oligopoly Model

Agents are companies competing in an oligopoly over multiple rounds:

Require:
Set of agents $\mathcal{N} = \{1, \dots, N\}$
Number of rounds T
For each agent i: initial demand D_1^i, unit production cost c_i, policy π_i, observation function Φ_i
for $t = 1, 2, \dots, T$ **do**
  for $i \in \mathcal{N}$ **do**
    i observes $o_t^i = \Phi_i(s_t) = \Phi_i(t, D_t^1, \dots, D_t^N)$
    i selects a price $p_t^i \sim \pi_i(o_t^i)$
    i sells quantity $D_t^i - p_t^i$
    i receives reward $r_t^i = (p_t^i - c_i)(D_t^i - p_t^i)$
  end for
  Compute the average price as $\bar{p}_t = \frac{1}{N} \sum_{j \in \mathcal{N}} p_t^j$
  for $i \in \mathcal{N}$ **do**
    Compute the price difference $\Delta p_t^i = p_t^i - \bar{p}_t$
    Transition demand to $D_{t+1}^i = D_t^i - \Delta p_t^i$
    Optionally, drop out i if $D_{t+1}^i < c_i$ (see Eq. (2))
  end for
end for
Reward each agent i with $U_i = \sum_{t=1}^{T} r_t^i$

The dropout mechanism removes unprofitable companies from the game

⟹ **Discontinuous game dynamics**

⟹ no analytical solutions with dropouts

Equilibrium Learning

Nash equilibria (NE) are strategy fixed points where no player gains by deviating:

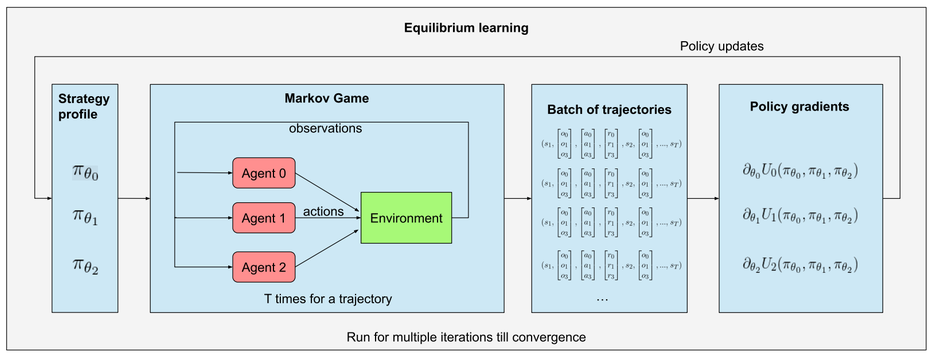
$$\sup_{\pi_i \in \Sigma_i} U_i(\pi_i, \pi_{-i}^*) - U_i(\pi_{\mathcal{N}}^*) \leq \varepsilon \quad \forall i \in \mathcal{N}$$

Proximity to equilibrium is measured by

$$\mathcal{L}_{\text{bf}} = \sum_{i \in \mathcal{N}} \sup_{\pi_i \in \Sigma_i^K} U_i(\pi_i, \pi_{-i}) - U_i(\pi_i, \pi_{-i})$$

where $\sup_{\pi_i \in \Sigma_i^K} U_i(\pi_i, \pi_{-i})$ is approximated by a brute force algorithm.

Each company is controlled by a RL agent aiming to maximize its profit:



We use REINFORCE (or its variants) to update all agents' policies simultaneously:

$$\theta_i \leftarrow \theta_i + \alpha \nabla_{\theta_i} U_i(\pi_{\theta_i}, \{\pi_{\theta_j}\}_{j \in \mathcal{N} \setminus \{i\}})$$

⟹ Hopefully converges to NE

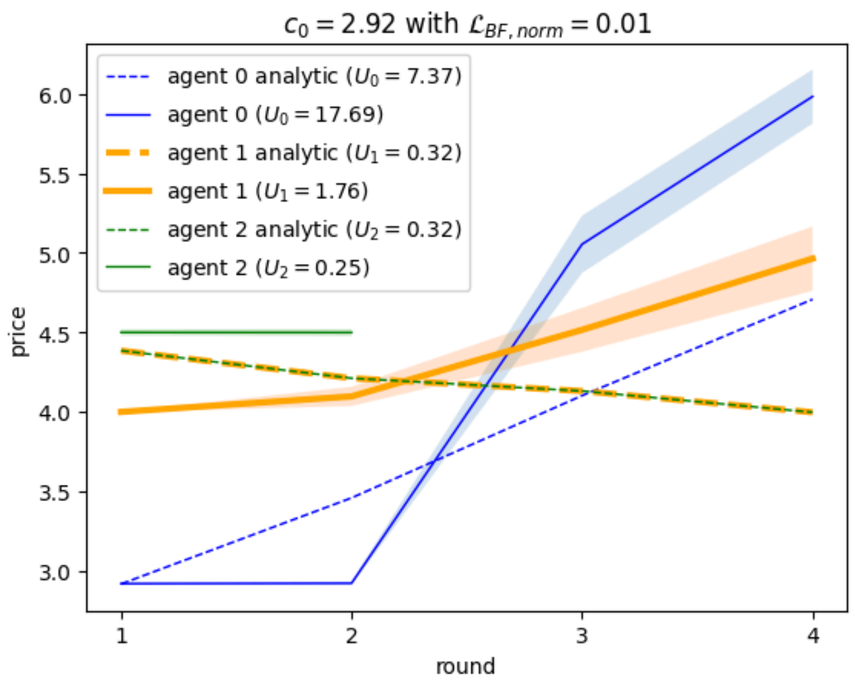
Results

Setup

- $N = 3$ companies, $T = 4$ rounds
- Proximal Policy Optimization (PPO) for RL
- Company 0 produces cheaper than 1 and 2

Findings

- Firm 0 learns **predatory pricing**, first lowering prices to drive out competitors
- The result is a verified approximate NE



Broader implications

- Our methodology finds approximate NE complex games without analytical solutions
- Brute-force verification is required due to lack of convergence guarantees in MARL

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Technische Universität München

zuseschoolrelai.de

relAI is supported by the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research.

SPONSORED BY THE Federal Ministry of Education and Research

DAAD Zuse Schools
Konrad Zuse Schools of Excellence in Artificial Intelligence