Rationality of Learning Algorithms in Repeated Games

Shivam Bajaj, Pranoy Das, Yevgeniy Vorobeychik, Vijay Gupta

1Electrical and Computer Engineering, Purdue University
2Computer Science and Engineering, Washington University in St. Louis

Introduction

- Game theory is a popular tool to model different types of interactions between multiple self-interested agents.
- Most learning algorithms designed for such agents in a game theoretic framework require every agent to adopt the same learning algorithm. Under this condition and for specific classes of games, these algorithms converge to an equilibrium, i.e., no agent has an incentive to deviate from the learning algorithm.
- An agent adopting a different learning algorithm may destabilize the system.
- A natural question is whether the agents have any incentive to adopt an alternative learning algorithm. If so, can we design algorithms in which an agent does not have any incentive to follow an alternative algorithm.

Problem Description

- Two player matrix games.
- Agents do not know the payoff matrix at start.
- Perfect monitoring assumption.
- A strategy profile \((\pi_1^*, \pi_2^*)\) is a Nash equilibrium if
  \[
  \mathcal{R}_1(\pi_1^*, \pi_2^*) \geq \mathcal{R}_1(\pi_1, \pi_2^*), \forall \pi_1 \neq \pi_1^*, \\
  \mathcal{R}_1(\pi_1^*, \pi_2^*) \geq \mathcal{R}_1(\pi_1, \pi_2), \forall \pi_1 \neq \pi_1^*.
  \]
- The value of player \(i\) is
  \[
  U_i(G, \mathcal{A}_1, \mathcal{A}_2) = \lim_{T \to \infty} \inf E \left[ \frac{1}{T} \sum_{t=0}^{T} \mathcal{R}_{i,t} \right]
  \]
- The rationality ratio of algorithm \(\mathcal{A}\) is
  \[
  s(\mathcal{A'}, \mathcal{A}) := \frac{U_i(\mathcal{A'}, \mathcal{A})}{U_i(\mathcal{A}, \mathcal{A})}.
  \]
- For a constant \(c \geq 1\), Algorithm \(\mathcal{A}\) is \(c\)-rational if \(\sup_{G, A'} s(\mathcal{A'}, \mathcal{A}) \leq c\).
- Algorithm \(\mathcal{A}\) is perfectly rational if \(c=1\).

Irrationality of Fictitious Play (FP)

Fictitious Play: Best respond to the empirical frequency of play of the other agent.

\[
BR_i(\hat{a}(-i(t-1))) := \arg \max_a \mathcal{R}_i(a, \hat{a}(-i(t-1)))
\]

Theorem: Algorithm FP is not \(c\)-rational for any \(c \geq 1\).

Rational Fictitious Play (R-GFP)

Exploration Phase: Agent \(i\) constructs \(E_i^t\) and selects action according to fictitious play. If agent \(-i\) deviates, agent \(i\) enters punishment phase.

Examples of \(E_i^t\) for \(t = 2\) and \(t = 3\).

Exploitation Phase: Agent \(i\) selects action according to fictitious play. If other agent deviates, agent \(i\) enters punishment phase.

Punishment Phase: Select actions that yield lowest payoff to agent \(-i\).

Theorem: Let \(\pi_1^*\) and \(\pi_2^*\) be the strategies of agent 1 and 2 if they selected actions according to fictitious play. Then,
- Algorithm R-GFP is perfectly rational if \(\min \max a_2 \mathcal{R}_{i,a_1} \leq U_i(\mathcal{A}, \mathcal{A})\)
- If \(\pi_1^*\) and \(\pi_2^*\) converge to Nash equilibrium, then so does Algorithm R-GFP.

Numerical Results

Future Work

Extension to multiple colluding agents, stochastic games, and imperfect monitoring scenarios.