ECE433/COS435 Spring 2024 Introduction to RL

Lecture 20: Game theory 101 and Multi-Agent RL

Game theory is a field of mathematics and economics that studies strategic interactions among rational decision-makers. It is also the theoretical foundation of multi-agent RL.

1. Basic Concepts

A game in the context of game theory is defined by three fundamental components:

- Players: The decision-makers in the game. Players can be individuals or entities.
- Strategies: The actions available to the players, from which they choose their course of action. A strategy can be pure, where a player selects a specific action, or mixed, where a player chooses a probability distribution over possible actions. In multi-agent RL, a pure strategy is a state-to-action policy map.
- Payoffs: The outcomes resulting from the combination of strategies chosen by all players, usually quantified in terms of utility or rewards.

In game theory, rationality assumes that each player will act to maximize their own payoff, given their knowledge of the strategies and payoffs involved.

1.1. Nash Equilibrium

The concept of Nash Equilibrium, named after mathematician John Nash, is central to game theory. It is a situation in which no player can benefit by changing their strategy while the other players keep theirs unchanged.

A strategy profile $s^* = (s_1^*, s_2^*, \ldots, s_n^*)$ is a Nash Equilibrium if for every player i,

$$u_i\left(s_i^*, s_{-i}^*\right) \geq u_i\left(s_i, s_{-i}^*\right)$$

for all s_i in the strategy space of player i, where u_i is the payoff function for player i and s_{-i}^* denotes the strategy profile of all players except i.

1.2. Two-Player Zero-Sum Game

In a two-player zero-sum game, each player aims to maximize their own payoff under the assumption that the other player is trying to minimize it. The interaction between the players can be represented as a matrix M, where each element $M(i, j)$ denotes the payoff to the first player (and thus a loss of the same magnitude to the second player) when the first player chooses action i and the second player chooses action j.

Minimax and Maximin Strategies

In this setup, player 1 and player 2 each choose a probability distribution over their respective actions. These distributions are represented by vectors p and q, where p and q are vectors in the simplex (i.e., their entries are non-negative and sum to 1).

- Maximin Strategy: Player 1 tries to maximize their minimum guaranteed payoff. Mathematically, this is expressed as:

$$\max_p \min_q p^T M q$$

- Minimax Strategy: Player 2 aims to minimize their maximum possible loss (which is equivalent to minimizing player 1's maximum payoff). This can be formulated as:

$$\min_q \max_p p^T M q$$

Strong Duality and Nash Equilibrium

In linear programming and convex optimization, strong duality holds when the primal and dual solutions converge to the same value. For the zero-sum game represented by $p^T Mq$, strong duality implies that the maximin and minimax values are equal:

$$\max_p \min_q p^T Mq = \min_q \max_p p^T Mq$$

This equality is crucial because it means that there is a saddle point in the game matrix M at (p^*, q^*), where p^* and q^* are optimal strategies for players 1 and 2 , respectively. This saddle point represents the **unique Nash equilibrium of any two-player zero-sum game**, where neither player can unilaterally change their strategy to improve their payoff given the strategy of the other player.

2. Learning in Games

To find the Nash equilibrium of games, we want to study dynamics that can learn to converge to the Nash.

2.1. Best Response Dynamics

The best response dynamics is a process in which players iteratively adjust their strategies to best respond to the strategies currently played by their opponents.

In a zero-sum game with payoff matrix M, let p_t and q_t represent the strategy profiles of players 1 and 2 , respectively, at time t. These strategies are probability distributions over their respective sets of actions. The best response dynamics is to iteration the following updates:

- $p_{t+1} = \arg\max_p p^T Mq_t$
- $q_{t+1} = \arg\min_q p_t^T Mq$

Players continue updating their strategies based on the best responses until no player can improve their payoff by unilaterally changing their strategy. However, it does not always converge to the Nash equilibrium.

Theorem: Best response dynamics does not converge for two-player zero-sum game.

Example: Rock-paper-scissor game.

2.2. Fictitious Play

- Consider a finite game that is played repeatedly in discrete time.
- The basic idea of fictitious play is that each player assumes that his opponents are using a fixed mixed strategy, and updates his beliefs about these strategies at each time step.
- Players choose actions in each time step to maximize that step's expected payoff given their belief of the opponents's strategies.
- The belief is formed as the empirical frequency distribution of the opponents's previously played strategies.

Player 1's strategy for the next period p_{t+1} is determined as the best response to Player 2's empirical distribution:

$$p_{t+1} = \arg\max_p p^T M(\frac{1}{t}\sum_{\tau=1}^{t} q_\tau)$$

Player 2's strategy for the next period p_{t+1} is determined as the best response to Player 1's empirical distribution:

$$q_{t+1} = \arg\min_q (\frac{1}{t}\sum_{\tau=1}^{t} q_\tau)^T Mq$$

Properties of fictitious play

- Myopic, because players are trying to maximize current payoff without considering their future payoffs. They are also not learning the "true model" generating the empirical frequencies (that is, how their opponent is actually playing the game).
- Not a unique rule due to multiple best responses. Traditional analysis assumes player chooses any of the pure best responses.

- Players do not need to know about their opponents's payoff; they only form beliefs about how their opponents will play.
- **Theorem:** Fictitious play converges to the unique Nash of a two-player zero-sum game.

3. Multi-Agent RL

The key technology of multi-agent RL is self-play, including playing against one-self, playing against previous versions of oneself, playing against other agents in the league. They are all various forms of best responses and fictitious play, rooted in game theory.

See link to google slides: https://docs.google.com/presentation/d/1aiWIFBYaBCJ_uqghXdgHmkdS89z1GZKy1OLpVlg_VC4/edit?usp=sharing