Deep Reinforcement Learning (Sp25)

Instructor: Dr. Mohammad Hossein Rohban

Quiz Solutions - Lecture 26

Arranged by: Behnia Soleymani

Q1

What is opponent modeling in MARL, and why is it often important for an agent's performance? Give an example of a simple opponent modeling technique.

Answer:
Opponent modeling in MARL is the process by which an agent tries to build a representation or prediction of other agents' behaviors, strategies, or intentions. It's important because an agent's optimal action often depends on what other agents are likely to do; better predictions about opponents can lead to better decision-making and higher rewards.

An example of a simple opponent modeling technique is Fictitious Play, where an agent keeps track of the historical frequency of actions taken by an opponent in specific states and uses these frequencies to predict the opponent's next action.

Q2

In the context of a general-sum stochastic game with N agents, which statement most accurately reflects the challenges and solution concepts?

A) There is always a unique, globally optimal Nash Equilibrium that all rational agents will converge to using standard Q-learning.

B) The game can be reliably solved by having each agent independently compute a min-max value function against all other N-1 agents combined.

C) Multiple Nash Equilibria often exist, and converging to a specific, socially desirable equilibrium is an open research problem; algorithms often focus on best-response dynamics or more specialized equilibrium concepts.

D) General-sum games always reduce to either fully cooperative or fully zero-sum games if agents are allowed to communicate their utility functions.

Correct Answer: C

Explanation: Options a, b, and d make overly strong or incorrect claims: a) uniqueness of NE and convergence via standard Q-learning is not guaranteed; b) min-max against all N-1 agents combined is not a general solution for general-sum games; d) general-sum games do not always reduce so simply even with communication.

Q3

Joint Q-Learning (JQL) is specifically designed for, and under ideal conditions (self-play, sufficient exploration, appropriate learning rate decay), converges to which type of equilibrium in what kind of stochastic game?

A) Min-max Nash equilibrium in competitive (zero-sum) stochastic games.
B) Pareto-dominating Nash equilibrium in cooperative stochastic games.
C) Any Nash equilibrium in general-sum stochastic games.
D) A correlated equilibrium in cooperative stochastic games.

Correct Answer: B
Explanation: Joint Q-Learning (JQL) is tailored for fully cooperative stochastic games. Under ideal conditions, it is designed to converge to a Nash equilibrium that is also Pareto-dominating.

Q4

In Minimax Q-Learning, when an agent j calculates the value of the next state s', denoted $V^j(s')$, for its Bellman update, it assumes:

A) All other agents will choose actions randomly based on a uniform distribution.
B) It will choose an action to maximize its Q-value, and all opponents will simultaneously choose their actions to minimize agent j's Q-value in state s'.
C) All other agents will continue to play according to their previously observed stationary mixed strategy.
D) It will choose an action to maximize its Q-value, and all other agents will also choose actions to maximize their own individual Q-values.

Correct Answer: B
Explanation: Minimax Q-Learning is designed for two-player zero-sum games. When agent j calculates $V^j(s')$, it assumes it will choose an action to maximize its Q-value, while the opponent will choose actions to minimize agent j's Q-value. This is the core of the minimax principle applied to the Q-learning update for the value of the next state.

Q5

Fictitious Play is an opponent modelling technique where an agent assumes that its opponents are:

A) Playing a stationary mixed strategy, which is estimated by counting their past actions.
B) Highly rational and will always play a min-max strategy against it.
C) Learning using the exact same algorithm as itself (self-play).
D) Intentionally trying to deceive it by appearing to play one strategy while planning another.

Correct Answer: A
Explanation: Fictitious Play operates on the assumption that opponents are playing a fixed (stationary) mixed strategy. The agent estimates this strategy by maintaining counts of the opponents' past actions and then plays a best response to this estimated empirical strategy.

Q6

Besides non-stationarity, describe two other significant challenges encountered in Multi-Agent Reinforcement Learning (MARL).

Answer: Besides non-stationarity, several other significant challenges are encountered in Multi-Agent Reinforcement Learning (MARL). Among them are:

- **Scalability (Growing Complexity):** The number of possible joint actions ($|A_1| \times \cdots \times |A_N|$) grows exponentially with the number of agents. Storing Q-values for all joint actions (as in naive joint Q-learning) or reasoning about this exponentially large action space becomes computationally infeasible very quickly.
- **Credit Assignment (Who Did What?):** In cooperative settings, especially when agents receive only a team reward, it is difficult to ascertain the individual contribution of each agent's action to the overall team performance. This makes it hard to effectively reinforce good individual behaviors or penalize detrimental ones.
- **Opponent Modelling (Guessing Others):** To act optimally, an agent often needs to predict or model the behavior, intentions, or strategies of other agents. This is challenging because other agents might also be learning and adapting, making their policies non-stationary from any single agent's perspective (this is related to, but distinct from, the general non-stationarity challenge).
- **Equilibrium Selection (Which NE?):** Many multi-agent games have multiple Nash Equilibria. If agents learn independently and converge to different, uncoordinated equilibria, the overall system performance can be poor. Ensuring convergence to a desirable (e.g., socially optimal or Pareto-efficient) equilibrium is a significant challenge.