Advanced Topics in Reinforcement Learning

Lecture 23: Multi-agent Learning I

Josiah Hanna
University of Wisconsin — Madison

Based on Slides from Stefano Albrecht's MARL lectures

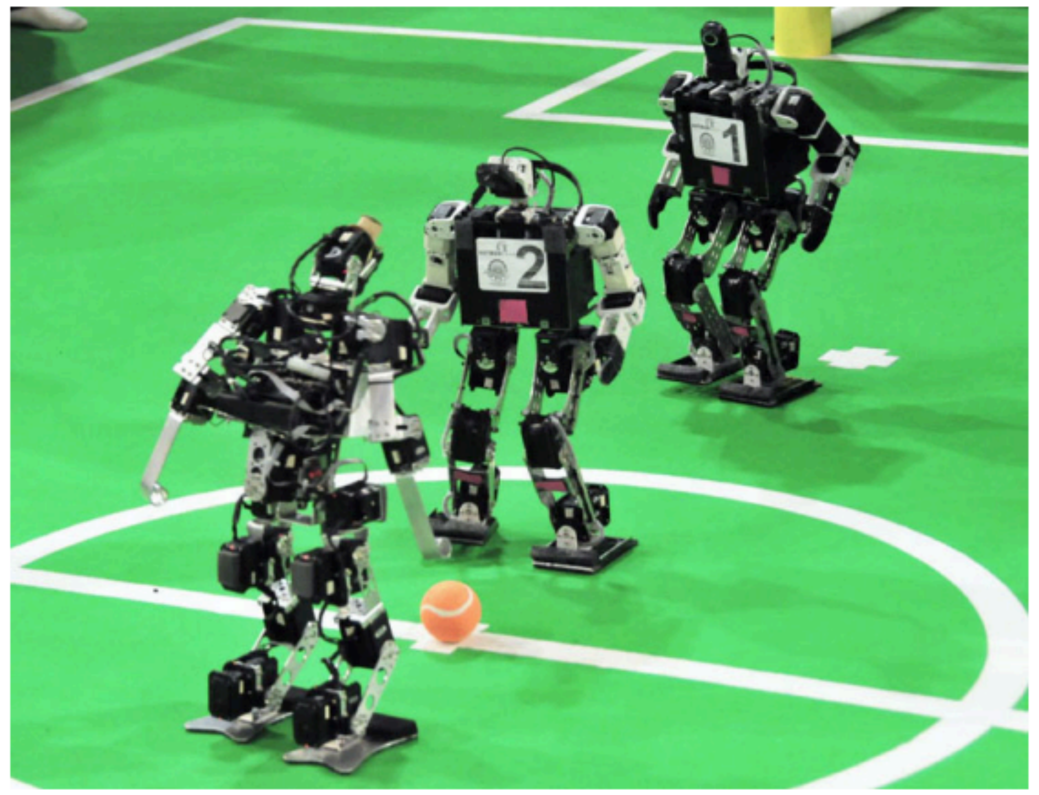
Announcements

- Next week: Offline RL
- Final projects due two weeks from Wednesday.
- Course evaluation is now available.

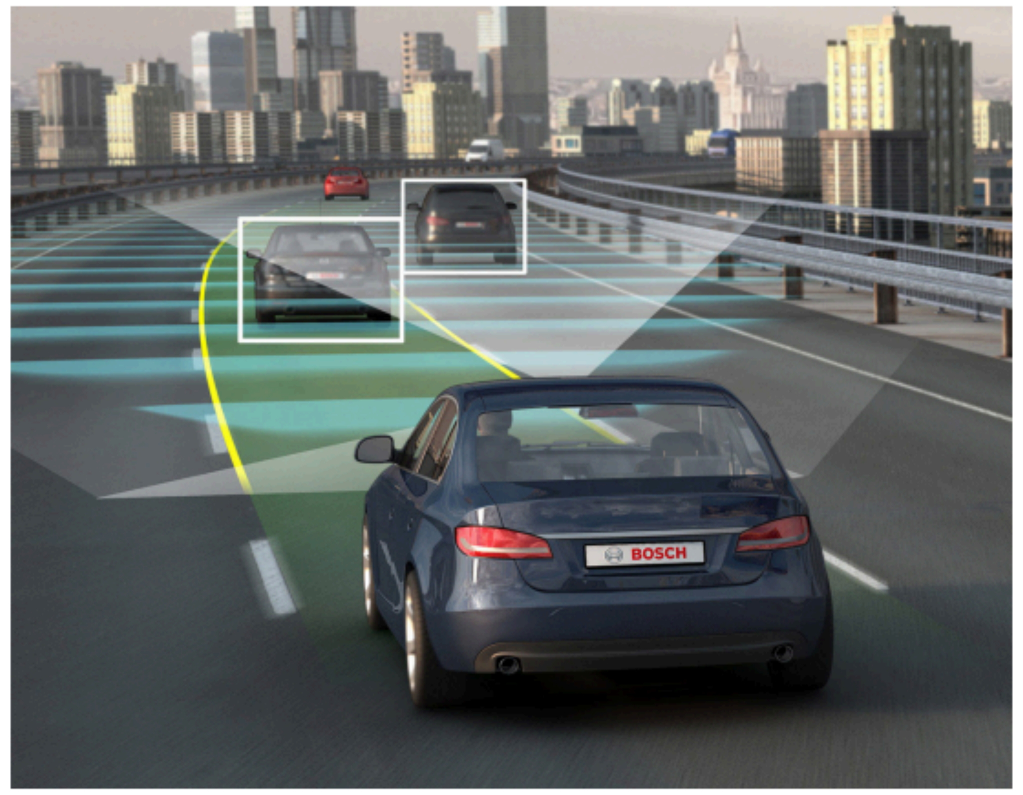
Multi-Agent Systems

Games

Robot soccer



Autonomous cars



Negotiation/markets

Wireless networks

Smart grid

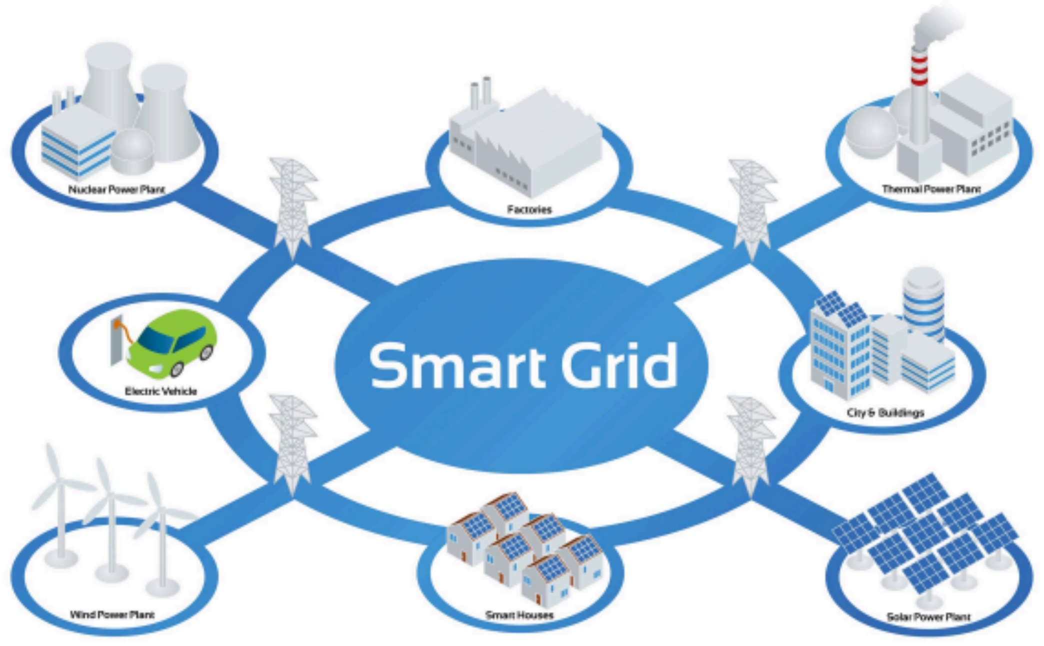


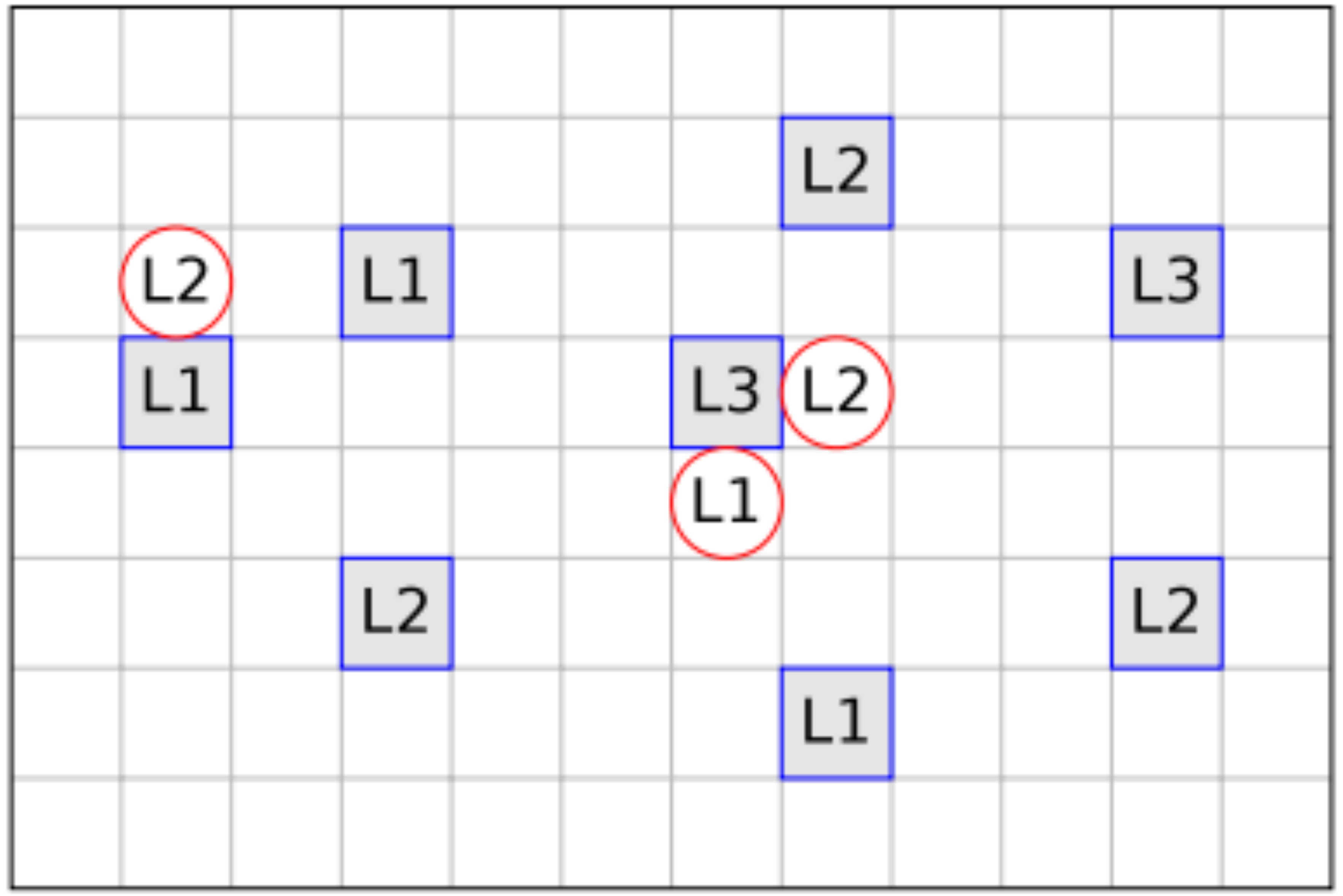
Image Credit: Stefano Albrecht

Challenges in Multi-Agent Learning

- Multi-agent credit assignment.
- Curse of multiple agents.
- Non-stationarity in learning.
- Different agents may have different objectives.

Multi-agent Credit Assignment

- All single-agent RL algorithms must solve temporal credit assignment.
  - Which actions contributed to eventual rewards received.
- Now each agent's rewards depend on what other agents do.
  - Did my action contribute when a reward was received?



Non-Stationarity in Multi-Agent Learning

- So far we have assumed that the environment's transition dynamics are stationary (unchanging over time).
- With learning, the environment appears non-stationary from the view of individual agents.
- Thus, the true action-values for any policy are also non-stationary.

Curse of Many Agents

- What if we just learn a policy that outputs an action for all agents?
  - Size of action space grows (possibly exponentially with number of agents).
  - Size of state space might grow.
  - Application communication constraints.
- Multi-agent RL decomposes a large RL problem into smaller, coupled problems.
- …but agents must coordinate action choices.

Stochastic Games

- Set of states $\mathcal{S}$.
- For each agent i:
  - Action set $\mathcal{A}_i$.
  - Reward function, $r_i : \mathcal{S} \times \mathcal{A}_1 \times \ldots \times \mathcal{A}_n \rightarrow \mathbf{R}$.
- Transition function, $p : \mathcal{S} \times \mathcal{A}_1 \times \ldots \times \mathcal{A}_n \times \mathcal{S} \rightarrow [0,1]$.
- Discount factor γ.

Interaction in Stochastic Games

- Begin in state s_0.
- At time t:
  - Each agent chooses action according to $\pi(A_t = a \mid S_t)$.
  - Each agent receives reward $r_i(S_t, A_t^1, \ldots A_t^n)$.
  - Transition to next state.
- How does this affect Markov property?

What do we want to converge to?

- Each agent wants to maximize reward but doing so depends on what other agents do.
  - Convergence defined in terms of policy profiles, $\pi = (\pi_1, \ldots, \pi_n)$.
- If all use the same reward function, then the optimal policy profile is to just maximize the expected return in each state.
- If not, many different solution concepts exist. Some examples:
  - Minimax optimality
  - Nash equilibrium
  - Pareto Optimality

Minimax Optimality

- A policy is minimax optimal for an agent if it has the best worst-case value.

- Typically considered in two player zero-sum games.

    - Two agents and $r_1(s, a_1, a_2) = -r_2(s, a_1, a_2)$.

- Agent 1 selects policy π; all other agents select the policy that makes π as bad as possible for Agent 1.

- Solution concept pursued in “Markov games as a framework for multi-agent reinforcement learning.”

Nash Equilibrium

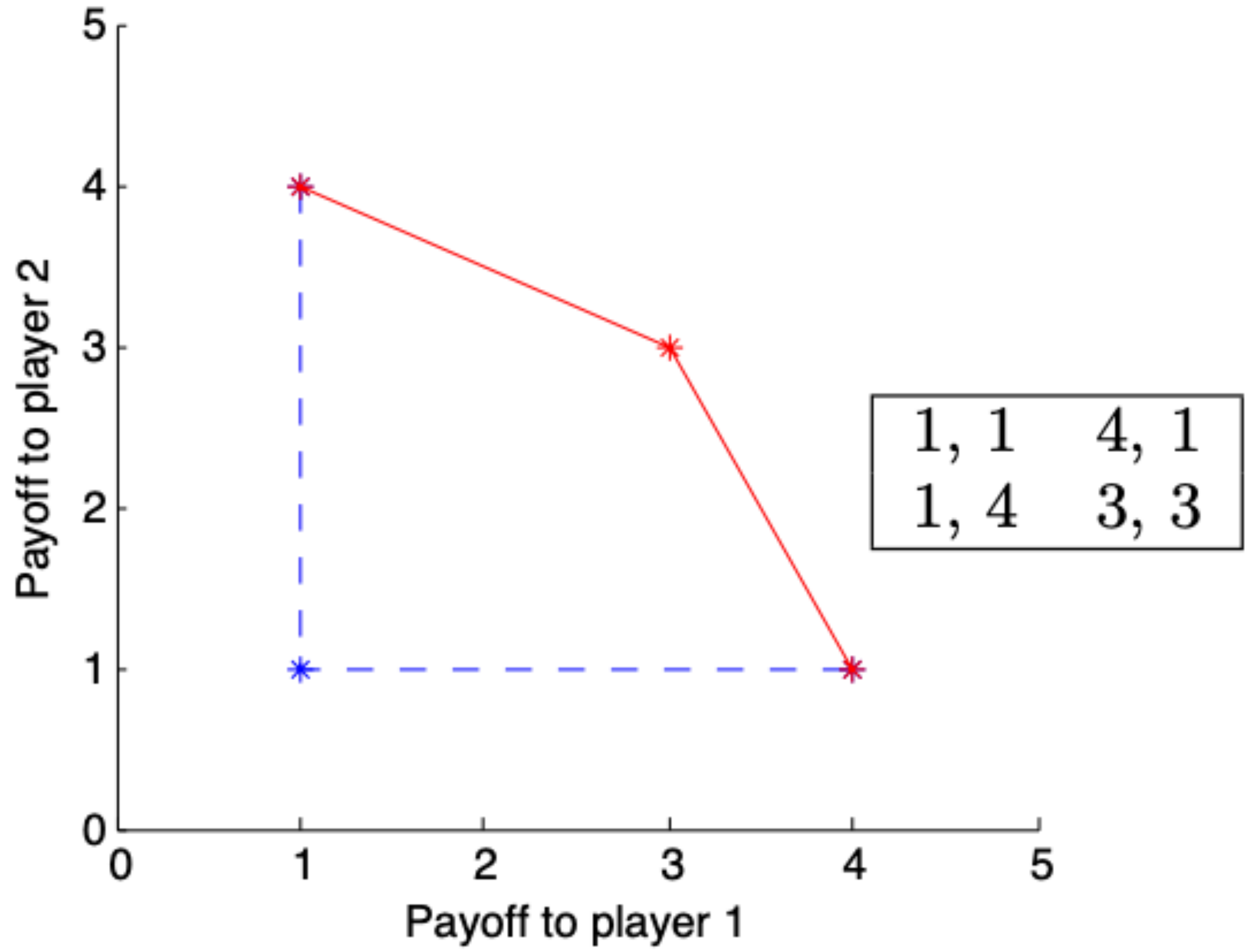
- A policy profile is a Nash equilibrium if no agent has an incentive to change their policy.
- Formally, profile π is a Nash equilibrium if $\forall i, \pi' \; v^i_{\pi'}(s) \leq v^i_{\pi}(s)$ where π' is identical to π except for agent i's policy.
- Assumes all agents are rational.

	C	D
C	-1,-1	-5,0
D	0,-5	-3,-3

Pareto Optimality

- Cannot improve one agent's value without decreasing another agent's value.

- Formally, a policy profile, π, is Pareto-optimal in state s if there is no other profile, π' such that $\forall i, v^i_{\pi'}(s) \geq v^i_\pi(s)$ and $\exists i, v^i_{\pi'}(s) > v^i_\pi(s)$.



Adam’s Presentation

- Slides

Summary

- Multi-agent RL aims to scale RL to environments with multiple, possibly learning agents.
- New challenges in MARL:
  - Credit assignment
  - Non-stationarity
- New solution concepts:
  - Minimax optimality, Pareto optimality, Nash equilibrium

Action Items

- Offline RL reading for next week.
- Good luck on your final project.