Move 37
Multi-Agent Reinforcement Learning - Study Guide

Basics (source)

- Merges game theory, machine learning, and cognitive science
- The state of the environment depends on the actions of every agent
- As the number of agents increases, the complexity greatly rises
- Multi-Agent RL suffers from the 'curse of dimensionality', a term Richard Bellman used to describe problems relating to organizing data in high dimensional spaces

Advantages (source)

- Robustness: failure of a single agent can be compensated by other members
- Scalability: more agents can be added over time
- Reusability: constituents can be retasked, while maintaining the system

Challenges

- Accomplishing global goals from local actions, each agent having limited observation
- Credit assignment: how do we know which agents contributed to successful trials?
- Incentives: how do we best reward our agents to accomplish our goal?
- Learning while others are learning: leads to a very dynamic (hard to predict) environment

Aspects

- Competition: increasing challenges between rivals lead to an automatic curriculum
  This principle underlies the success of AlphaGo
- Cooperation: includes learning from others
- Communication: includes negotiation
- Social Dilemmas: can be useful for studying emergent cooperation and competition (these are often borrowed from game theory)

Reality-import

- We live in a multi-agent world
- An organization can be thought of as a multi-agent architecture
- Traffic, the economy, and the environment are all multi-agent systems
- AI assistant: an ideal digital assistant would make accurate predictions about a human's mental state and intentions (it would have a good theory of mind).
- AI safety and ethics usually involves the interplay of humans and intelligent systems.

Please watch this video when you have a chance. The speaker is a great authority on AlphaGo and move 37.

Inverse Reinforcement Learning - Study Guide

Basics (source)

- Instead of using the rewards to find the policy, as we do in normal RL, Inverse RL seeks to find a reward function based on a given policy (behavior observed).
- We assume that the expert demonstrator is acting optimally (vanilla IRL)
- Inverse RL is sometimes called inverse optimal control (IOC)
- 'Algorithms for Inverse Reinforcement Learning' (2000) is the original paper that lays the foundation for IRL (source)

Use cases (source)

- Modeling animal (including human) behavior
- Apprenticeship Learning/Learning from Demonstration
- Modeling of agents for cooperation/competition ('Theory of Mind' for multi-agent RL)
- This area of research has potential ethical consequence for how we interact with our technology (related to theory of mind) (source)

Inverse Reinforcement Learning (IRL) vs Imitation Learning (IL) (source)

- Both IRL and IL are forms of Learning from Demonstration (LfD)
- In IRL we find a reward function based on expert demonstration
- In IL we try to generalize an expert strategy to unvisited states (similar to classification)
- It's possible to represent both using a unified framework, which achieves better results than either alone
- The term 'behavior cloning' is sometimes used interchangeably with imitation learning

Maximum Likelihood Inverse Reinforcement Learning (source)

1. Guess a reward function
2. Find a policy that's optimal for this reward function
3. Get the probability of the demonstration data, given this policy
4. Adjust the reward function in the direction of a gradient so as to increase the likelihood of the demonstration data
   (repeat steps 2-4 until…)
5. Stop when we reach a local maximum

Check out the following papers on theory of mind (not on the test):

- Machine Theory of Mind (2018)
- Modelling User's Theory of AI's Mind in Interactive Intelligent Systems (2018)
- M^3RL: Mind-aware Multi-agent Management Reinforcement Learning (2018)
- Intrinsic Social Motivation via Causal Influence in Multi-Agent RL (2018)