INSTITUTO SUPERIOR TÉCNICO

Multiagent models for partially observable environments

Matthijs Spaan

Institute for Systems and Robotics

Instituto Superior Técnico

Lisbon, Portugal

Reading group meeting, March 26, 2007

Overview

- Multiagent models for partially observable environments:
  - Non-communicative models.
  - Communicative models.
  - Game-theoretic models.
  - Some algorithms.
- Talk based on survey by Frans Oliehoek (2006).

The Dec-Tiger problem

- A toy problem: decentralized tiger (Nair et al., 2003).
- Two agents, two doors.
- Opening correct door: both receive treasure.
- Opening wrong door: both get attacked by a tiger.
- Agents can open a door, or listen.
- Two noisy observations: hear tiger left or right.
- Don’t know the other’s actions or observations.

Multiagent planning frameworks

Aspects:

- communication
- on-line vs. off-line
- centralized vs. distributed
- cooperative vs. self-interested
- observability
- factored reward

Partially observable stochastic games

Partially observable stochastic games (POSGs) (Hansen et al., 2004):

- Extension of stochastic games (Shapley, 1953).
- Hence self-interested.
- Agents do not observe each other's observations or actions.

POSGs: definition

- A set $I = \{1, \ldots, n\}$ of n agents.
- A_i is the set of actions for agent i.
- O_i is the set of observations for agent i.
- Transition model $p(s'|s, \bar{a})$ where $\bar{a} \in A_1 \times \ldots \times A_n$.
- Observation model $p(\bar{o}|s, \bar{a})$ where $\bar{o} \in O_1 \times \ldots \times O_n$.
- Reward function $R_i : S \times A_1 \times \ldots \times A_n \rightarrow \mathbb{R}$.
- Each agents maximizes $E\Big[\sum_{t=0}^{h} \gamma^t R_i^t\Big]$.
- Policy $\pi = \{\pi_1, \ldots, \pi_n\}$, with $\pi_i : \times_{t-1}(A_i \times O_i) \rightarrow A_i$.

Decentralized POMDPs

Decentralized partially observable Markov decision processes (Dec-POMDPs) (Bernstein et al., 2002):

- Cooperative version of POSGs.
- Only one reward, i.e., reward functions are identical for each agent.
- Reward function $R : S \times A_1 \times \ldots \times A_n \to \mathbb{R}$.

Dec-MDPs:

- Jointly observable Dec-POMDP: joint observation $\bar{o} = \{o_1, \ldots, o_n\}$ identifies the state.
- But each agents only observes o_i.

MTDP (Pynadath and Tambe, 2002): essentially identical to Dec-POMDP.

Interactive POMDPs

Interactive POMDPs (Gmytrasiewicz and Doshi, 2005):

- For self-interested agents.
- Each agents keeps a belief over world states and other agents' models.
- An agent's model: local observation history, policy, observation function.
- Leads to infinite hierarchy of beliefs.

Communication

- Implicit or explicit.
- Implicit communication can be modeled in "non-communicative" frameworks.
- Explicit communication Goldman and Zilberstein (2004):
  - informative messages
  - commitments
  - rewards/punishments
- Semantics:
  - Fixed: optimize joint policy given semantics.
  - General case: optimize meanings as well.
- Potential assumptions: instantaneous, noise-free, broadcast communication.

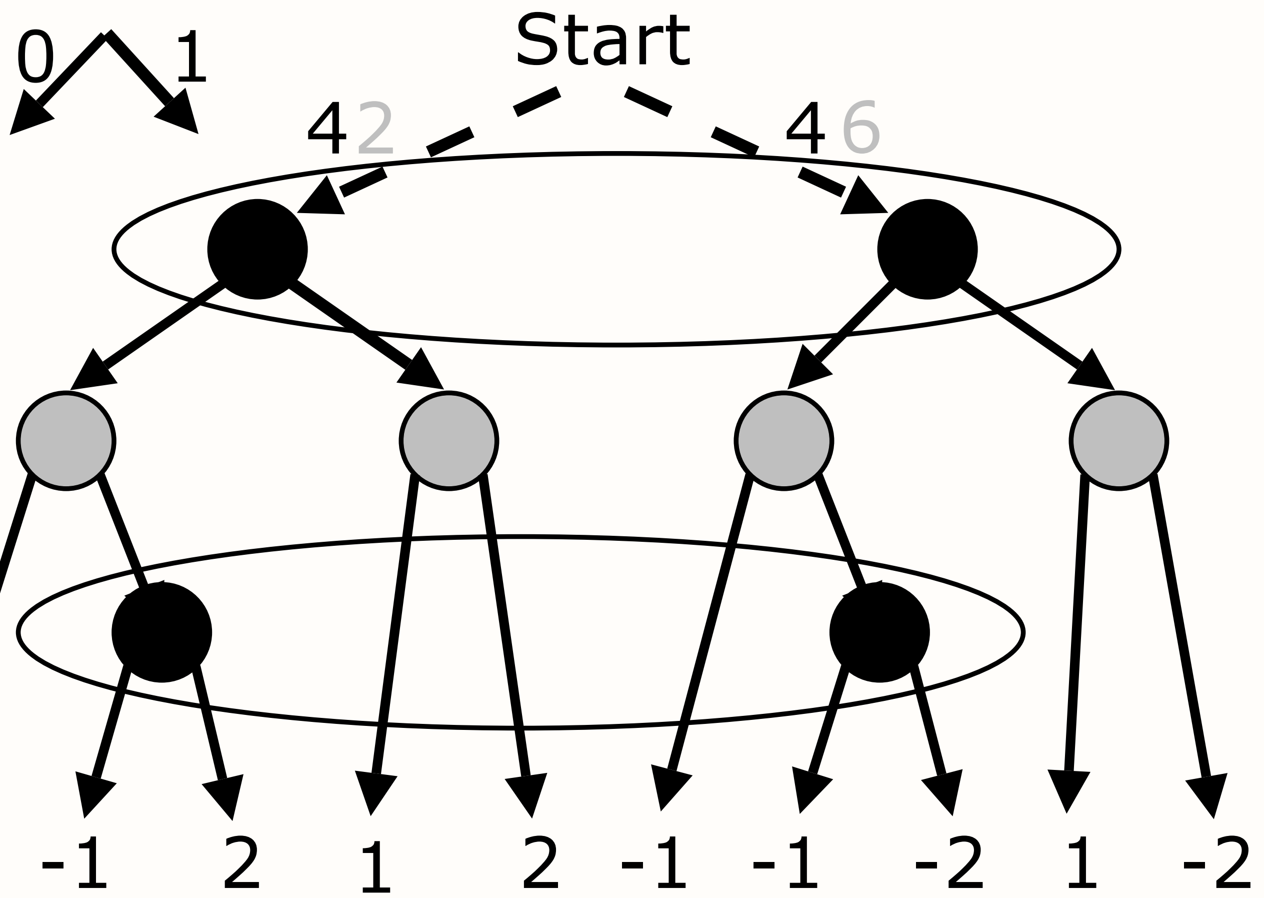
Dec-POMDPs with communication

Dec-POMDP-Com (Goldman and Zilberstein, 2004)

- Dec-POMDP plus:
- Σ is the alphabet of all possible messages.
- σ_i is a message sent by agent i.
- $C_\Sigma : \Sigma \rightarrow \mathbb{R}$ is the cost of sending a message.
- Reward depends on message sent: $R(s, a_1, \sigma_1, \ldots, a_n, \sigma_n, s')$.
- Instantaneous broadcast communication.
- Fixed semantics.
- Two policies: for domain-level actions, and for communicating.
- Closely related model: Com-MTDP (Pynadath and Tambe, 2002).

Extensive form games

8-card poker:



Extensive form games (1)

Extensive form games:

- View a POSG as a game tree.
- Agents act on information sets.
- Actions are taken in turns.
- POSGs are defined over world states, extensive form games over nodes in the game tree.

Dec-POMDP complexity results

	Observability			
Communication	fully	jointly	partial	none
none	P	NEXP	NEXP	NP
general	P	NEXP	NEXP	NP
free, instantaneous	P	P	PSPACE	NP

Dynamic programming for POSGs

- Dynamic programming for POSGs (Hansen et al., 2004).
- Uncertainty over state and the other agent's future conditional plans.
- Define value function V_t over state and other agent's depth-t policy trees: a $|S|$ vector for each pair of policy trees.
- Computing the $t + 1$ value function requires backing up all combinations of all agents' depth-t policy trees.

  $\Rightarrow$ Prune (very weakly) dominated strategies.
- Optimal for cooperative settings (DEC-POMDP).
- Still infeasible for all but the smallest problems.

(Approximate) DEC-POMDP solving

- Extra assumptions: e.g., independent observations, factored state representation, local full observability (DEC-MDP), structure in the reward function.
- Optimize one agent while keeping others fixed, and iterate.
  $\Rightarrow$ Settle for locally optimal solutions.
- Free communication turns problem into a big POMDP.
  $\Rightarrow$ Find good on-line communication policy.
- Add synchronization action (Nair et al., 2004).
- Belief over belief tree (Roth et al., 2005).

Some algorithms

Joint Equilibrium based Search for Policies (Nair et al., 2003)

- Use alternating maximization.
- Converges to Nash equilibrium, which is a local optimum.
- Keeps belief over state and other agents' observation histories.
- This POMDP is transformed to an MDP over the belief states, and solved using value iteration.

Some algorithms (1)

Set-Coverage algorithm Becker et al. (2004):

- For transition-independent Dec-MDPs with a particular joint reward structure.

Bounded Policy Iteration for Dec-POMDPs (Bernstein et al., 2005):

- Optimize a finite-state controller with a bounded size.
- Alternating maximization.

References

R. Becker, S. Zilberstein, V. Lesser, and C. V. Goldman. Solving transition independent decentralized Markov decision processes. *Journal of Artificial Intelligence Research*, 22:423–455, 2004.

D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein. The complexity of decentralized control of Markov decision processes. *Mathematics of Operations Research*, 27(4):819–840, 2002.

D. S. Bernstein, E. A. Hansen, and S. Zilberstein. Bounded policy iteration for decentralized POMDPs. In *Proc. Int. Joint Conf. on Artificial Intelligence*, 2005.

P. J. Gmytrasiewicz and P. Doshi. A framework for sequential planning in multi-agent settings. *Journal of Artificial Intelligence Research*, 24:49–79, 2005.

C. V. Goldman and S. Zilberstein. Decentralized control of cooperative systems: Categorization and complexity analysis. *Journal of Artificial Intelligence Research*, 22:143–174, 2004.

E. A. Hansen, D. Bernstein, and S. Zilberstein. Dynamic programming for partially observable stochastic games. In *Proc. of the National Conference on Artificial Intelligence*, 2004.

R. Nair, M. Tambe, M. Yokoo, D. Pynadath, and S. Marsella. Taming decentralized POMDPs: Towards efficient policy computation for multiagent settings. In *Proc. Int. Joint Conf. on Artificial Intelligence*, 2003.

R. Nair, M. Tambe, M. Roth, and M. Yokoo. Communication for improving policy computation in distributed POMDPs. In *Proc. of Int. Joint Conference on Autonomous Agents and Multi Agent Systems*, 2004.

D. V. Pynadath and M. Tambe. The communicative multiagent team decision problem: Analyzing teamwork theories and models. *Journal of Artificial Intelligence Research*, 16:389–423, 2002.

M. Roth, R. Simmons, and M. Veloso. Decentralized communication strategies for coordinated multi-agent policies. In A. Schultz, L. Parker, and F. Schneider, editors, *Multi-Robot Systems: From Swarms to Intelligent Automata*, volume IV. Kluwer Academic Publishers, 2005.

L. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39:1095–1100, 1953.