Learning Probabilistic Models via Bayesian Inverse Planning

Abdeslam Boularias and Brahim Chaib-draa

Laval University, Canada

{boularias;chaib}@damas.ift.ulaval.ca

DAMAS Décision, Adaptation, Multi-AgentS
www.damas.ift.ulaval.ca

UNIVERSITÉ LAVAL

1. Overview

- Inverse planning refers to the problem of learning a probabilistic model of the dynamics of a robot (or a human) by observing her optimal behavior performing a given task.
- This work extends previous work on inverse reinforcement learning [Ng et Russell, 2000, Ramachandran & Amir, 2007, Baker et al., 2009] to the problem of learning a transition function when a reward function is given.
- We show that inverse planning can be efficiently used for cooperative planning in teams of heterogenous robots.

2. Model

A Transition-Independent Multi-Agent Markov Decision Process (MMDP) is defined by:

- $\mathcal{I}$: a set of agents.
- $\{\mathcal{S}^i\}$: a set of states per agent.
- $\{\mathcal{A}^i\}$: a set of actions per agent.
- $\{T^i\}$: a set of independent transition functions, where $T^i(s,a,s')$ is the probability that agent i will end up in state s' after taking action a in state s.
- R: a reward function, $R(s,a)$ is the reward that all the agents receive when they execute the joint action a in joint state s.

3. Problem

Assumptions:

1. Every agent i knows only her own transition function T^i.
2. The agents know the reward function R.
3. The agents can observe the states and actions of the others.
4. The agents cannot communicate.

Problem: Find an optimal joint policy under these assumptions.

4. A Toy Example

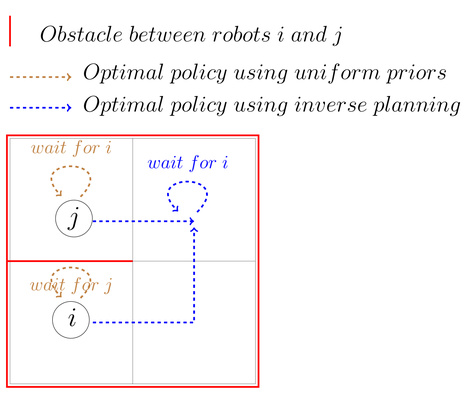
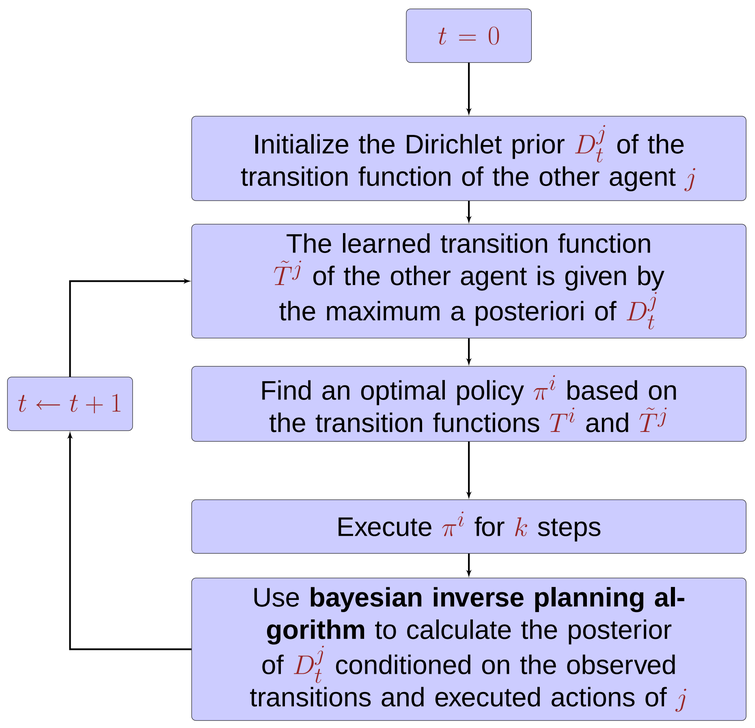


Figure 1: *Meeting problem 1.*

- Robots i and j want to meet, but they cannot communicate.
- Actions of both i and j succeed with probability 0.7
- Neither i nor j can cross the obstacle.
- Using a uniform prior about the transition function of the other robot (a 0.5 probability to cross the obstacle), the optimal policy is to stay and wait for the other robot to cross the obstacle.
- By observing the other robot waiting, and using bayesian inverse planning, every robot learns that the probability that the other one can cross the obstacle is less 0.5

5. Planning

The planning algorithm executed by an agent i:



6. Bayesian Inverse Planning

The posterior distribution of a transition function T given a prior D_t and a trajectory $H_t = \{(s_0,a_0,s_1),\ldots,(s_k,a_k,s_{k+1})\}$ is:

$$\underbrace{Pr(T|D_t,H_t)}_{\text{posterior distribution}} \propto \prod_{(s,a,s')\in H_t} \underbrace{Pr(s'|s,a,T)}_{\text{transition evidence}} \underbrace{Pr(a|s,T)}_{\text{policy evidence}} \underbrace{Pr(T|D_t)}_{\text{prior}}$$

- We assume that the actions are given by a softmax function:

$$Pr(a|s,T) \propto e^{\alpha_t Q_T^*(s,a)}$$

$Q_T^*(s,a)$ are calculated by using the transition function T.

- The posterior distribution does not have a closed form
- We use a gradient ascent algorithm to calculate a local maximum a posteriori of T.

7. Example II

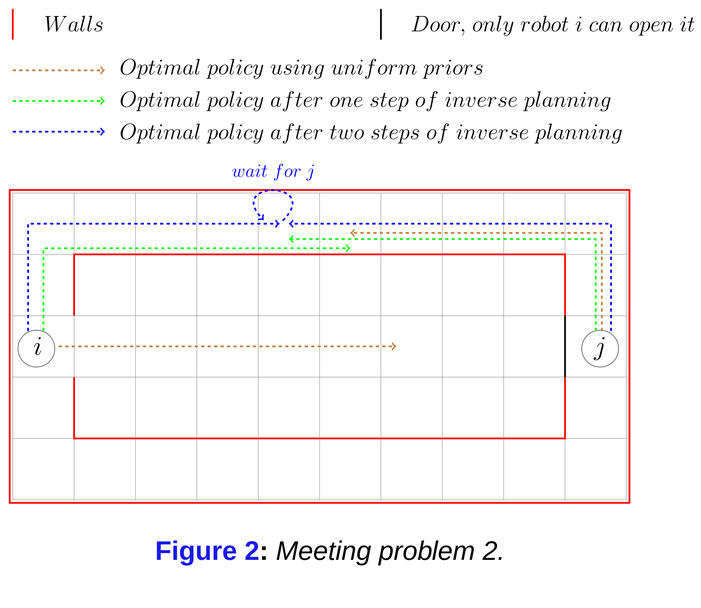
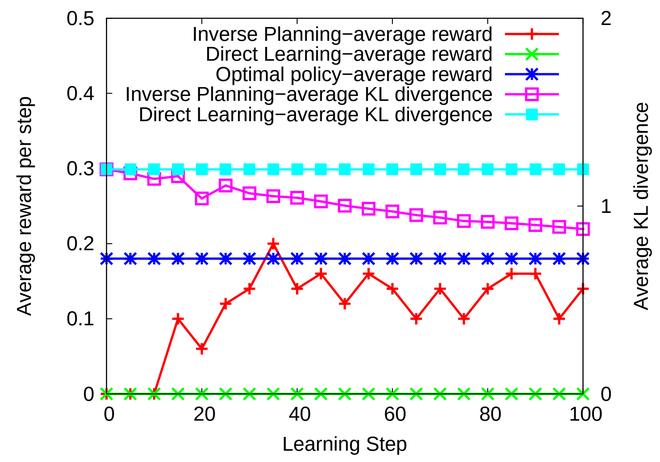


Figure 2: *Meeting problem 2.*

8. Preliminary Results



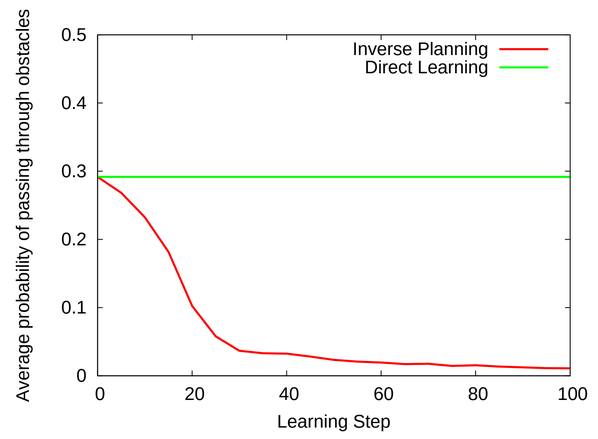


Figure 3: *Results on the meeting problem 1.*

9. Conclusion and Future Work

- ✓ Inverse planning can be used to learn the parameters related to states that are not even contained in the collected data.
- ✓ Preliminary results show that this technique is promising for cooperative planning problems.
- ✗ The complexity of calculating the gradient is $O(|\mathcal{S}|^4)$.
- ✗ The results of the gradient heavily depend on the stepsize.
- ✗ The tradeoff between the transition evidence and the policy evidence is not well-captured by the vanilla gradient.

As a future work, we target to:

- Find an appropriate riemannian metric and use the natural gradient.
- Find a fast approximation of the gradient.
- Use inverse planning in tasks involving a human-machine interaction.

References

[Ng et Russell, 2000] Ng, A., & Russell, S. (2000). Algorithms for Inverse Reinforcement Learning. *Proceedings of the Seventeenth International Conference on Machine Learning (ICML'00)* (pp. 663–670).

[Ramachandran & Amir, 2007] Ramachandran, D., & Amir, E. (2007). Bayesian Inverse Reinforcement Learning. *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI'07)* (pp. 2586–2591).

[Baker et al., 2009] Baker, C., Saxe, R., & Tenenbaum, J. (2009). Action Understanding as Inverse Planning. Cognition, 113, (pp. 329–349).