Learning to Plan Under Variable Assistance

Rohan Chitnis*, Jessica Yu*, Lilian Luong, Leslie Pack Kaelbling, Tomás Lozano-Pérez

MIT Computer Science and Artificial Intelligence Laboratory
{ronuchit, yujess, luong, lpk, tlp}@mit.edu

I. Introduction

Modern household and industrial robots have demonstrated proficiency in executing repetitive physical tasks when acting alone or with other robots. Work in human-robot collaboration typically assumes a consistent human presence [1], [2], [3], [4], and often frames the problem as a human-led enterprise. However, in real-world settings such as households and factories, a human manager might only occasionally be available to assist a robot with its task.

We term such settings *variably assistive*: an autonomous robot intermittently receives assistance from another agent (e.g., a human) on a task it is performing. These environments are by design partially observable, in the sense that the robot must estimate aspects of the assistive agent's availability, in order to plan around the times that they will be there to help with the task. For instance, a factory robot transporting boxes of varying weights should reason about when assistance is expected to be available, and accordingly form plans that will move heavy boxes only while assisted.

Acting optimally in variably assistive settings requires the robot to learn characteristics of the assistive agent from experience, such as their schedule and their general eagerness to help. Furthermore, the robot must balance expectations with the risk of investing time into multi-step subtasks predicated on hypothesized projections of the availability.

In this work, we motivate a framework for describing variably assistive settings using the notion of *exogenous processes*, which have typically been studied in the context of rule-based planning [5], [6], [7], [8]. We experiment with a simulated discrete planning scenario in which a robot must transport boxes, and the presence of the assistive agent endows it with additional carrying capacity. Our results show that explicitly estimating future availability of the assistive agent yields better task performance than do a set of baseline heuristics making various assumptions about the availability.

II. Modeling Variable Assistance Problems

The central idea in variably assistive settings is that the availability of the assistive agent is outside the robot's control, but needs to be estimated. From the robot's perspective, then, this availability can be viewed as an *exogenous process*, unaffected by the robot's own actions. Formally, an exogenous process $\{x_t\}$ is one whose dynamics satisfy

$$P(x_{t+1} \mid x_t, a_t; \theta) = P(x_{t+1} \mid x_t; \theta),$$

where a_t is the robot's action at time t, and the θ are any parameters of the dynamics model.

To model intermittent assistance, we propose to introduce an exogenous process that defines a latent "countdown" until the next toggle in availability. The countdown itself is a stochastic process driven by latent parameters θ that can encode characteristics of the assistive agent.

Let $s_t \in \mathbb{R}^d$ describe the state of the environment and robot, which includes, for instance, object poses and the robot configuration. For clarity, we present our model under the assumption that s_t is fully observed, though generalizing it to partially observed s_t is straightforward.

We define the *variable assistance* problem as a partially observable Markov decision process with state space $\mathcal{S}$, action space $\mathcal{A}$, observation space Ω, and reward function R. A state in $\mathcal{S} = \mathbb{R}^{d+m} \times (\mathbb{N} \setminus \{0\})$ and an observation in $\Omega = \{0, 1\} \times \mathbb{N}$ can be written as:

$$\begin{bmatrix} s_t & c_t & \theta \end{bmatrix} \in \mathcal{S}, \qquad \begin{bmatrix} H_t & e_t \end{bmatrix} \in \Omega.$$

Here, s_t is observed (we omit it from Ω for clarity), and:

- $H_t \in \{0, 1\}$ is an observed indicator for whether the assistance is currently available.
- $c_t \in \mathbb{N} \setminus \{0\}$ represents a latent countdown until the next toggle in H_t will occur, which will also reset c_t.
- $\theta \in \mathbb{R}^m$ are a static set of latent parameters governing the dynamics of what value c_t resets to.
- $e_t \in \mathbb{N}$ is the observed time since the last toggle in H_t.

We make the assumption that the $\{c_t\}$ are an exogenous process, and so we can factor the model by making certain conditional independence assumptions, shown in Figure 1.

Since c_t and θ are unobserved by the robot, it must maintain a belief state, a probability distribution over their values conditioned on the history of observations:

$$B_t = P(c_t, \theta \mid H_0, e_0, ..., H_t, e_t).$$

The latent counter c_t serves to make the process Markov: it would be unrealistic to assume that the availability indicator H_t depends only on its predecessor H_{t-1}, but when conditioned on c_t, the Markov assumption $P(H_t \mid H_0, H_1, ..., H_{t-1}, c_t) = P(H_t \mid H_{t-1}, c_t)$ is sensible.

We define the models needed for belief state estimation:

$$P(c_{t+1} \mid c_t, \theta) = \begin{cases} \mathbb{1}[c_t - c_{t+1} = 1] & c_t > 1 \\ f_\theta(c_{t+1}) & c_t = 1 \end{cases}$$

[*] Equal contribution.

Presented at the ICRA Workshop on Robot Teammates Operating in Dynamic, Unstructured Environments (RT-DUNE), 2019.

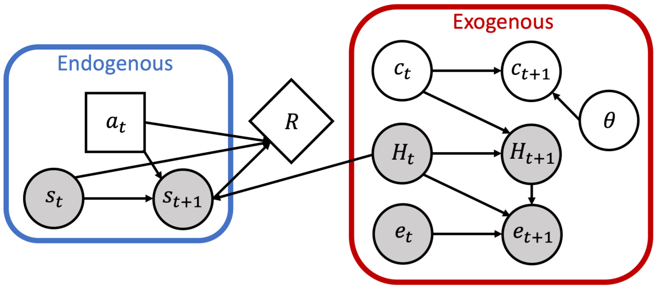

Fig. 1: Our model of the variable assistance problem setting. We view the presence of the assistive agent as an exogenous process from the robot's perspective, in the sense that the robot's actions do not affect this process. In order to plan, the robot must estimate the latent state of the process, since it affects the state transitions. In this diagram, shaded nodes represent observed variables.

$$P(H_{t+1} \mid H_t, c_t) = \begin{cases} \mathbb{1}[H_{t+1} = H_t] & c_t > 1 \\ \mathbb{1}[H_{t+1} = 1 - H_t] & c_t = 1 \end{cases}$$

$$P(e_{t+1} \mid e_t, H_t, H_{t+1}) = \begin{cases} \mathbb{1}[e_{t+1} - e_t = 1] & H_t = H_{t+1} \\ \mathbb{1}[e_{t+1} = 0] & H_t \neq H_{t+1} \end{cases}$$

Here, f_θ is a distribution over $\mathbb{N} \setminus \{0\}$ parameterized by θ, representing possible new values of the counter each time it is reset (which also triggers a toggle in the availability H_t). Doing belief state estimation with these models leads to learning about characteristics of the assistive agent, described in the marginal distribution over θ implied by the belief.

The optimal solution to a variable assistance problem is a policy, a mapping from beliefs to actions, that maximizes the expected reward over the trajectory: $\mathbb{E}[\sum_t R(s_t, a_t, s_{t+1})]$. The dependence of s_{t+1} on H_t will be task-specific: it captures how the robot is affected by the exogenous process.

One could imagine a simple extension of this model that allows f_θ to depend on the states of other exogenous processes. For instance, a human may be more generally available to assist their household robot on weekends rather than weekdays, and so f_θ could depend on the current day of the week. For clarity, we have described only the simplest version of the model here, with one exogenous process.

III. Preliminary Results

We experiment with a simulated discrete planning scenario in which a robot must transport boxes, and the presence of the assistive agent endows it with additional carrying capacity. We consider two settings of f_θ: a deterministic one in which $c_{t+1} = \theta$, and the prior on θ is uniform; and a stochastic one in which $c_{t+1} \sim \text{Poisson}(\theta)$, and the prior on θ is a gamma distribution. We perform state estimation using a variant of particle filtering that resamples particles from an analytically maintained posterior on θ (leveraging the conjugate priors) whenever the filter collapses, for robustness to situations where the true state is underrepresented within the particle set. We plan in belief space using the maximum likelihood observation assumption [9], and replan whenever the optimistic assumptions are violated during execution.

As baselines, we consider three heuristic methods that plan according to various assumptions about the dynamics of H_t, instead of estimating c_t or θ. "Reactive" assumes that the currently observed H_t will hold forever; "Greedy" greedily tackles the most difficult subtask possible under the currently observed H_t; "Delay" tackles subtasks in increasing order of difficulty, disregarding the current availability.

System	Avg. Cost: Deterministic	Avg. Cost: Stochastic
Oracle	26.2	26.1
Reactive	38.1	36.94
Greedy	38.5	38.12
Delay	44.2	39.55
AdaptEx	**27.6**	**33.31**

TABLE I: Average execution costs on box transport task, over 20 maps and 5 random seeds per map. "Oracle" is an oracle robot that can observe c_t and θ. AdaptEx is our estimation and planning system that reasons about and adapts to exogenous processes.

Table I shows execution costs averaged across 100 random initial object configurations. We can see that reasoning about the exogenous process to estimate future availability of the assistive agent yields better task performance than do the baseline approaches, especially for deterministic settings.

IV. Next Steps

We aim to expand the ideas presented here toward a general framework for reasoning about partially observed exogenous processes in planning. If the robot can identify regularities in the dependency structure among these processes, it could leverage these patterns to factor its belief, making inference more tractable. In addition, the robot could learn low-dimensional embeddings of the exogenous process states that capture only the aspects relevant to the problem it is currently facing, thus speeding up planning considerably.

References

[1] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, "Effects of robot motion on human-robot collaboration," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 51–58.

[2] D. Hadfield-Menell, S. J. Russell, P. Abbeel, and A. Dragan, "Cooperative inverse reinforcement learning," in *Advances in neural information processing systems*, 2016, pp. 3909–3917.

[3] C. Morato, K. N. Kaipa, B. Zhao, and S. K. Gupta, "Toward safe human robot collaboration by using multiple Kinects based real-time human tracking," *Journal of Computing and Information Science in Engineering*, vol. 14, no. 1, p. 011006, 2014.

[4] Y. Li and S. S. Ge, "Human-robot collaboration based on motion intention estimation," *IEEE/ASME Transactions on Mechatronics*, vol. 19, no. 3, pp. 1007–1014, 2014.

[5] G. De Giacomo, R. Reiter, and M. Soutchanski, "Execution monitoring of high-level robot programs," in *Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann Publishers, 1998.

[6] M. Beetz and D. V. McDermott, "Improving robot plans during their execution." in *AIPS*, 1994, pp. 7–12.

[7] A. Gerevini, A. Saetti, and I. Serina, "An approach to temporal planning and scheduling in domains with predictable exogenous events," *Journal of Artificial Intelligence Research*, vol. 25, pp. 187–231, 2006.

[8] L. Iocchi, D. Nardi, and R. Rosati, "Planning with sensing, concurrency, and exogenous events: logical framework and implementation," in *KR*. Citeseer, 2000, pp. 678–689.

[9] R. Platt Jr., R. Tedrake, L. Kaelbling, and T. Lozano-Perez, "Belief space planning assuming maximum likelihood observations," 2010.