Estimating Value of Assistance for Online POMDP Robotic Agents

Yuval Goshen and Sarah Keren
Taub Faculty of Computer Science, Technion - Israel Institute of Technology
Emails: yuval.goshen@campus.technion.ac.il sarahk@cs.technion.ac.il

***Abstract*—Robotic agents are increasingly required to operate in dynamic, uncertain, and partially observable environments. Our goal is to develop effective multi-robot systems that can make principled real-time decisions about when and how to assist their teammates. We support multi-robot settings in which an agent can offer assistance to a team member by modifying the physical environment, for example by moving obstacles that block the ability to perceive or reach objects. These assistance capabilities create opportunities for meaningful collaboration in multi-robot systems but introduce a challenging decision-making problem: assuming the ability to assist is limited, what is the most effective helping action among the multiple alternatives, and, in case there are multiple agents that may require assistance, which agent should be assisted next?**

To address these challenges, we offer a principled approach for quantifying and comparing the potential benefits of assistive actions by formulating *Value of Assistance* (VOA) for robotic POMDP agents, and by developing computationally efficient heuristics for VOA estimation. Our empirical evaluation on both a standard POMDP benchmark and a collaborative manipulation setting demonstrates how our suggested measures enable real-time decision-making while maintaining sufficient accuracy for helping action selection.

I. Introduction

Robotic agents are required to operate in increasingly dynamic and unpredictable environments, in which they lack complete information about the world due to sensor noise, a limited field of view, or incomplete models of the environment. In this work, we support multi-robot settings in which agents can assist each other in different ways. For example, a team member can move obstacles or rearrange objects to facilitate manipulation or to improve observability.

These assistance options create opportunities for meaningful collaboration in multi-robot systems but introduce several decision-making challenges; An agent must select the most effective helping action from multiple alternatives, evaluate whether pausing its current task to assist another agent would be worthwhile, and determine which agent would benefit most from assistance when multiple team members may require assistance.

To address these challenges, our work focuses on formulating ways for estimating *Value of Assistance* (VOA) for partially informed robotic agents and on providing principled approaches for quantifying and comparing the potential benefits of different assistive actions.

Example: [1] *To demonstrate, consider our two-agent evaluation setting depicted in Figure 1[left]. Agent1 on the right is equipped with a parallel gripper and a camera, and is tasked with a complex manipulation task: placing two cups on the table and pouring soda into them from a can. For this, the agent must plan and execute high-level actions - such as picking and placing objects - while moving to positions that enable manipulation or sensing of the workspace.*

In this collaborative setting, Agent2 on the left can offer assistance by moving obstacles (which is not possible for Agent1) using its vacuum gripper, though it has its own task. Figure 1[right] presents two possible options. On the top, moving Obstacle #1 enables observation of the previously occluded can. On the bottom, moving Obstacle #2 facilitates grasping the blue cup. We aim to support the decision of the assisting robot by proposing ways to quantify Value of assistance (VOA) as the expected long-term benefit of an assistive action. This will allow selecting the best action and comparing its expected benefit against the cost incurred from pausing Agent2's task.

Beyond this illustrative example, assessing VOA is relevant to a broader class of applications such as automated manufacturing, environmental monitoring, and construction in which agents can assist each other but need to consider their own resources and objectives. In such settings, deciding how to assist other agents requires considering the long-term effects of actions. For example, will clearing a path benefit only the current navigation segment or enable access to multiple future goals?

As is typical in such settings, we model the agent's task using a Partially Observable Markov Decision Process (POMDPs) [1]. POMDPs provide a mathematical framework for decision-making under uncertainty and support principled probabilistic reasoning, balancing the trade-off between exploring to gather more information and exploiting current knowledge to maximize reward. POMDPs are particularly relevant to the robotic settings we aim to support since they allow capturing the complexities of noisy sensors, imperfect actuation, and partial observability inherent in real-world robotic tasks. However, together with their expressive power, exact POMDP solutions are computationally intractable for large problems [6]. This has led to the development of

[1]The accompanying video demonstrates this setup and assistance scenario in action: https://drive.google.com/file/d/19xzithddSJAobt-dG6F0aQVZpKnTjqwp/view?usp=sharing

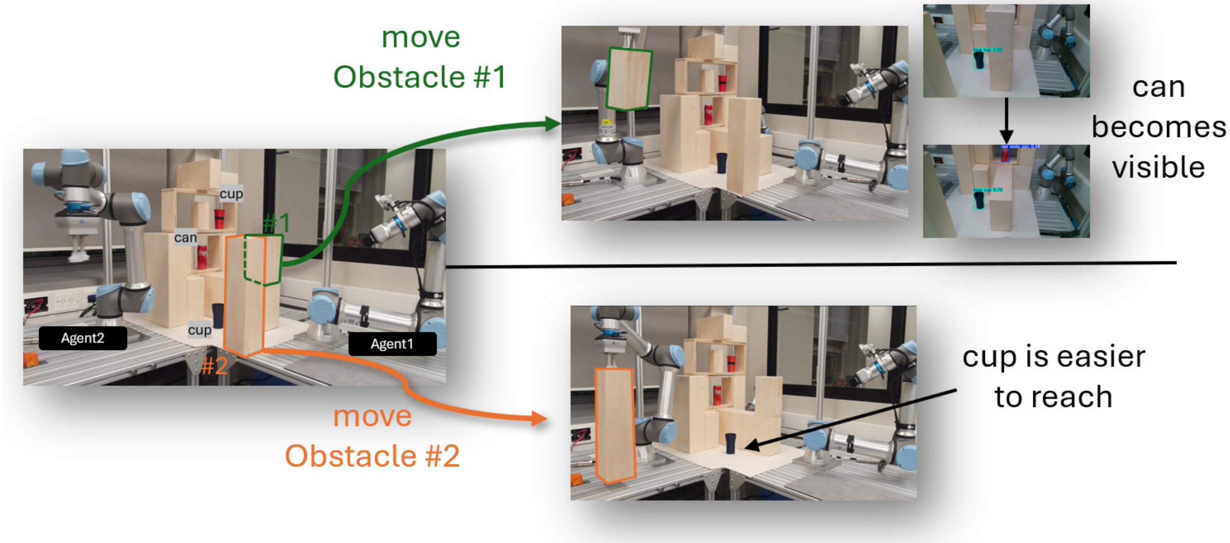

Fig. 1: A collaborative multi-robot setting. [left] Initial setup: Agent1 has a camera and a parallel gripper that can only manipulate the smaller objects in the scene, and needs to pour soda from the can into the two cups, after placing them on the table. Agent2 has a Vaccum gripper with which it can pick up the large wooden blocks. It might have its own task, but can pause it to assist Agent1 by moving one of the larger blocks. [right] Two assistance options: on the top, by moving Obstacle#1, the previously occluded can become visible. On the bottom, moving Obstacle#2 makes it easier to reach the blue cup. We formulate VOA measures to compare the expected benefit of each assistance action and choose the best one.

various online planning methods that have enabled practical applications in robotics [3, 2].

As a result of the complexity of POMDP decision making and its online, approximate nature, exact computation of VOA becomes intractable for real-time decision making, especially when there are many possible interventions to consider and there is a need reevaluate the policy to determine the effect interventions will have the agent's decisions.

With the objective of supporting effective real-time assistance decisions, we formulate Value of Assistance (VOA) for POMDP agents and address computational challenges by developing domain-agnostic heuristics that enable rapid evaluation of potential helping actions while accounting for their long-term effects.

We formulate Value of Assistance (VOA) as the difference between the expected future rewards an agent would accumulate with and without assistance. For POMDP agents using online planning, directly computing this value through policy evaluation is computationally prohibitive for real-time decision-making. To address this challenge, we develop three domain-agnostic heuristics for efficiently approximating VOA:

- **First-Action Value** (h_{FA}): Approximates VOA by planning only the first step with and without assistance, using the MCTS search tree's root node value estimate.
- **Rollout-Policy** ($h_{\pi_{Rollout}}$): Uses the planner's internal rollout policy directly for action selection, avoiding the computational cost of tree construction.
- **Full-Information** (h_{FO}): Transforms the POMDP into a fully-observable deterministic planning problem, inspired by [4, 5], evaluating assistance based on optimal deterministic plans.

We evaluated these heuristics on both a modified RockSample POMDP benchmark and a collaborative robotic manipulation task. Our empirical results demonstrate that the Full-Information heuristic provides the best balance of computational efficiency and accuracy across domains. With computation times under 0.1 seconds, it achieves high correlation with empirically measured VOA values and consistently identifies beneficial helping actions. While the First-Action Value heuristic also demonstrates promising results with modest computational requirements, the Rollout-Policy approach proves inadequate for capturing the complex long-term effects of assistance actions. These findings enable real-time decision-making about assistance in multi-robot systems while maintaining sufficient accuracy for effective helping action selection.

References

[1] Anthony R Cassandra, Leslie Pack Kaelbling, and Michael L Littman. Acting optimally in partially observable stochastic domains. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 1023–1028, 1994.

[2] Hanna Kurniawati. Partially observable markov decision processes and robotics. *Annual Review of Control, Robotics, and Autonomous Systems*, 5(Volume 5, 2022):253–277, 2022. ISSN 2573-5144. doi: https://doi.org/10.1146/annurev-control-042920-092451. URL https://www.annualreviews.org/content/journals/10.1146/annurev-control-042920-092451.

[3] Mikko Lauri, David Hsu, and Joni Pajarinen. Partially observable markov decision processes in robotics: A survey. *IEEE Transactions on Robotics*, 39(1): 21–40, February 2023. ISSN 1941-0468. doi: 10.1109/tro.2022.3200138. URL http://dx.doi.org/10.1109/TRO.2022.3200138.

[4] Mausam and Andrey Kolobov. *Planning with Markov Decision Processes: An AI Perspective*. Morgan & Claypool Publishers, 2012. ISBN 1608458865.

[5] Andrey Kolobov Mausam and Daniel S. Weld. Determinize, solve, and generalize: Classical planning for mdp heuristics. 2009. URL https://api.semanticscholar.org/CorpusID:13457525.

[6] Christos H. Papadimitriou and John N. Tsitsiklis. The complexity of markov decision processes. *Mathematics of Operations Research*, 12(3):441–450, 1987. ISSN 0364765X, 15265471. URL http://www.jstor.org/stable/3689975.