Planning and Reinforcement Learning for General-Purpose Service Robots

Yuqian Jiang
University of Texas at Austin, Texas, US
jiangyuqian@utexas.edu

Abstract

Despite recent progress in AI and robotics research, especially learned robot skills, there remain significant challenges in building robust, scalable, and general-purpose systems for service robots. This Ph.D. research aims to combine symbolic planning and reinforcement learning to reason about high-level robot tasks and adapt to the real world. We will introduce task planning algorithms that adapt to the environment and other agents, as well as reinforcement learning methods that are practical for service robot systems. Taken together, this work will make a significant step towards creating general-purpose service robots.

1 Introduction

It has been a longstanding goal of AI and robotics research to create intelligent robots that are competent at solving problems, adaptable in complex environments, and able to naturally interact with humans. In particular, general-purpose service robots have been envisioned to operate ubiquitously in homes and public spaces, carrying out a wide range of tasks such as cooking, cleaning and delivery.

However, there remains a significant gap between specialized robot skills that work well in constrained lab conditions and integrated robots that act robustly in the open world [Kunze *et al.*, 2018]. Despite deep learning being largely responsible for the performance improvements in low-level robot skills, learning alone is not sufficient for the high-level task control of service robots. Consider a non-expert user's request "get me an apple from the kitchen". In order to accomplish this task robustly, a robot has to plan for the command, possibly ask for clarifications, search probable locations of apples, and explain if an apple cannot be retrieved. Such high-level reasoning problems are beyond the capabilities of current neural networks due to issues such as sample efficiency, reliability, and explainability.

In order to build high-level control of general-purpose robots, symbolic AI methods can be leveraged to provide useful abstractions and guarantees. Given a definition of the domain, symbolic planning has long been able to decompose complex goals into symbolic actions that are then executed with lower-level controllers, enabling dynamic sequencing of individual robot skills. While classical planning methods generate provably feasible plans with respect to the model of the environment available at planning time, they are often brittle if there is missing information, or if the environment changes over time. The uncertain, dynamic, and human-populated real world therefore calls for an integration of planning with learning to adapt to the environment. One promising learning approach is reinforcement learning (RL), which allows for autonomous agents to learn optimal behavior through interactions with the environment.

To close the gap between specialized skills and general-purpose robots, this Ph.D. research will investigate the following question: *How can symbolic planning and RL be combined to create general-purpose service robots that reason about high-level actions and adapt to the real world?* This extended abstract will introduce our initial contributions towards this goal.

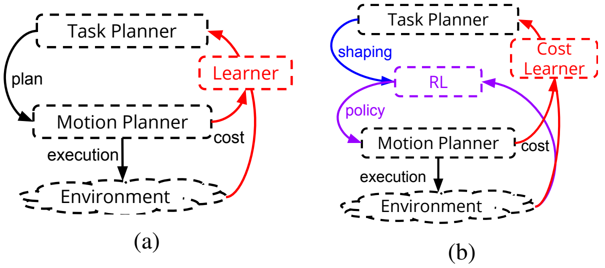
2 Planning

Symbolic planning aims at helping an agent plan a sequence of actions to accomplish complex tasks given a definition of action preconditions and effects. Motivated by challenges in enabling general-purpose service robots [Walker *et al.*, 2019], this work will make the following planning contributions:

Open-world reasoning and planning. A service robot accepting verbal commands from a human operator is likely to encounter requests that reference objects not currently represented in its knowledge base. We have introduced a method that reasons about hypothetical objects from interactions with human users in order to plan in the open world [Jiang *et al.*, 2019a].

Multi-robot planning with conflicts and synergies. In multi-robot environments, robots encounter situations where they have conflicting needs for constrained resources, or where they can take advantage of what each other is doing to form synergies. This work formulates the problem of multi-robot planning with conflicts and synergies (MRPCS), and develops a multi-robot planning framework, called *iterative inter-dependent planning* (IIDP), for representing and solving MRPCS problems [Jiang *et al.*, 2019c].

Task-motion planning with cost learning. Task and motion planning for a mobile service robot can be sensitive to domain uncertainty and changes, leading to suboptimal behaviors or



(a)

(b)

Figure 1: (a) TMP-CL framework (b) Proposed framework of integrating RL for action selection.

execution failures. This work proposes the *TMP-CL* framework (shown in Figure 1a), which generates a low-cost, feasible task-motion plan by iteratively planning and updating the relevant action costs evaluated by the motion planner. During execution, the framework learns actual action costs to further improve its task-motion plans [Jiang *et al.*, 2019b].

3 Reinforcement Learning

Applying RL to physical robots often suffers from delayed or sparse rewards, slowing down learning significantly due to long stretches of possibly uninformative exploration. Much of RL research today focuses on the episodic setting with discounted total reward, which is suitable for learning individual robotic skills. But continuing tasks, which are common in long-term deployments of robots, often need to maximize some form of long-term average reward [Mahadevan, 1996]. This work will develop methods for average-reward RL, and leverage reasoning and planning to speed up training.

Continual area sweeping. A service robot must update its knowledge of the environment that it operates in. We formalize the problem of a service robot surveying its environment non-uniformly as maximizing the average reward in a Semi-Markov Decision Process, and introduce a deep RL algorithm for such a setting [Shah *et al.*, 2020].

Temporal-logic-based reward shaping for average-reward RL. Reward shaping [Ng *et al.*, 1999] is a common approach for incorporating domain knowledge into RL to speed up learning. We introduce the first reward shaping framework for average-reward learning and prove that, under standard assumptions, the optimal policy under the original reward function can be recovered. In order to avoid the need for manual construction of the shaping function, we introduce a method for automatically translating a temporal-logic formula to a shaping function [Jiang *et al.*, 2021].

Reinforcement learning for task-motion planning (future work). The *TMP-CL* framework introduced above learns action costs in execution but ignores unexpected action outcomes. Since the action definitions are manually specified for the symbolic planner, the transitions cannot be modified easily based on execution experiences. In future work, we plan to explore the idea of integrating model-free RL for action selection as shown in Figure 1a. The high-quality initial plans and fast convergence of *TMP-CL* can still be leveraged via reward shaping.

4 Conclusion

Taken together, these contributions will be integrated and evaluated on real robots, leading to a concrete realized general-purpose service robot: such a robot can solve daily tasks such as delivering objects and tidying, while being robust to missing or inaccurate information in the model and adaptive to the real world. When all requests are serviced, the robot carries out the idle task of continual area sweeping to keep its knowledge up to date. In cases where multiple robots operate in the same space, they are able to adapt the high-level plans to avoid collisions and form synergies.

References

[Jiang *et al.*, 2019a] Yuqian Jiang, Nick Walker, Justin Hart, and Peter Stone. Open-world reasoning for service robots. In *Proceedings of the 29th International Conference on Automated Planning and Scheduling (ICAPS 2019)*, July 2019.

[Jiang *et al.*, 2019b] Yuqian Jiang, Fangkai Yang, Shiqi Zhang, and Peter Stone. Task-motion planning with reinforcement learning for adaptable mobile service robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2019)*, November 2019.

[Jiang *et al.*, 2019c] Yuqian Jiang, Harel Yedidsion, Shiqi Zhang, Guni Sharon, and Peter Stone. Multi-robot planning with conflicts and synergies. *Autonomous Robots*, March 2019.

[Jiang *et al.*, 2021] Yuqian Jiang, Suda Bharadwaj, Bo Wu, Rishi Shah, Ufuk Topcu, and Peter Stone. Temporal-logic-based reward shaping for continuing reinforcement learning tasks. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI 2021)*, February 2021.

[Kunze *et al.*, 2018] Lars Kunze, Nick Hawes, Tom Duckett, Marc Hanheide, and Tomáš Krajník. Artificial intelligence for long-term robot autonomy: A survey. *IEEE Robotics and Automation Letters*, 3(4):4023–4030, 2018.

[Mahadevan, 1996] Sridhar Mahadevan. Average reward reinforcement learning: Foundations, algorithms, and empirical results. *Machine learning*, 22(1-3):159–195, 1996.

[Ng *et al.*, 1999] Andrew Y Ng, Daishi Harada, and Stuart J Russell. Policy invariance under reward transformations: Theory and application to reward shaping. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pages 278–287. Morgan Kaufmann Publishers Inc., 1999.

[Shah *et al.*, 2020] Rishi Shah, Yuqian Jiang, Justin Hart, and Peter Stone. Deep r-learning for continual area sweeping. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2020)*, October 2020.

[Walker *et al.*, 2019] Nick Walker, Yuqian Jiang, Maya Cakmak, and Peter Stone. Desiderata for planning systems in general-purpose service robots. In *Proceedings of the ICAPS Workshop on Planning and Robotics (PlanRob 2019)*, July 2019.