# Constructing Policies for Supportive Behaviors and Communicative Actions in Human-Robot Teaming

Elena Corina Grigore
Yale University, Department of Computer Science
elena.corina.grigore@yale.edu

Brian Scassellati
Yale University, Department of Computer Science
brian.scassellati@yale.edu

***Abstract*—Current state-of-the-art robotic systems deployed in industry work in isolation from humans and do not allow for collaboration. Developing a robot that can work side-by-side with a human presents the advantage of allowing both the robot and the human worker to focus on the task each is best suited for, while assisting one another as needed. For the robot to provide assistive behavior to a human co-worker, it needs to learn what actions it should perform at each time step depending upon the state of the task. Such assistive actions are not intended to simply contribute to the completion of a particular task by instructing the robot to work on subtasks in isolation from the human worker; rather they are meant to help the worker complete the task more efficiently. As such, employing standard policy search or task and motion planning techniques is not sufficient to discover the supportive types of actions my system seeks to offer based on accurate estimations of the current task state. To this end, my research focuses on investigating policy search within hierarchical tasks that allow for two main abilities, namely helping the human co-worker more effectively complete a task and taking communicative actions that reduce state estimation uncertainty by asking the worker direct questions. The policy dictates what action the robot should take at each time step, based on inputs from a motion capture system providing observations about the configuration of the person's hands relative to the objects needed for accomplishing the task, as well as the person's answers to any questions posed by the robot.**

## I. Introduction

Robotics research today aims to push the limits of robot autonomous capabilities. Whether it be robots that help people in their homes, public spaces, hospitals, or assembly-line settings, their ability to perform assistive behaviors is of paramount importance. Given that most systems currently deployed in industry are not equipped to work side-by-side with humans, a big challenge is developing robots capable of offering assistance to human workers for a variety of tasks.

Assistive behaviors help workers more efficiently complete a task (e.g. a person would assemble a chair frame faster if the robot could stabilize the frame). Standard task and motion planning techniques based on configuration spaces that employ a divide-and-conquer approach cannot identify such assistive behaviors. This is because the necessary actions are not distinct components of the task given to the system a priori but need to be discovered based on each subtask performed. Developing new policy search algorithms is thus necessary. The focus of the presented research topic is to find policies that dictate what action the robot should perform to both provide assistive behavior and reduce its uncertainty about current task state.

## II. Background and Related Work

The current work employs a reinforcement learning (RL) framework, allowing the specification of a reward function for learning. RL problems are framed based on Markov Decision Processes (MDPs), which describe the environment where an agent can act. The aim is to learn a policy that chooses actions such that it locally optimizes the delayed cumulative reward signal, defined based on the objective of the task. Typically, this is achieved by parameterizing the policy indirectly through estimating state or action values (value-approximation methods) or through directly parameterized policies (direct policy search methods). While reviews encompassing these methods [1], [2] already highlight the challenges faced by typical robotics tasks that work with continuous state and/or action spaces, the problem at hand requires the adaptation of such standard techniques and poses greater challenges still.

To tackle the problem presented, I use hierarchical RL [3], [4], dividing the task into subtasks for which I apply different types of policy search. At the low level, the policies handle continuous state spaces and strive to find actions that help implement the assistive behaviors. At the high level the policies handle uncertainty due to states not fully observable (extending the MDP to the partially observable variant POMDP) and strive to find actions that either decide what kind of supportive behavior to provide or reduce state estimation uncertainty.

In the human-robot interaction field, work that focuses on human-robot teaming includes research in the area of search and rescue [5], team coordination behaviors and action planning for human-robot teaming [6], and dynamic sharing of responsibilities between a robot and a human operator [7]. Going beyond the important aspects investigated in related work, my research focuses on including two capabilities into the action policy: (1) the ability to model supportive behaviors that do not necessarily directly contribute to task completion but help the human worker more efficiently complete the task at hand, and (2) allow the robot to deliberately act so as to reduce uncertainty about state estimation by taking communicative actions (asking the user direct questions).

## III. Approach

In this work I use a typical RL MDP formulation. An MDP is a tuple $(S, A, P, R, \gamma)$, where $S$ represents the state space the agent operates in, $A$ represents the action space the agent acts within, $P$ is a probabilistic transition function such

978-1-4673-8370-7/16/$31.00 © 2016 IEEE

that $P(s, a, s') = P(s_{t+1}|s_t = s \text{ and } a_t = a)$ denotes the probability that action $a$ in state $s_t$ results in a transition to state $s_{t+1}$, $R$ represents the reward function that specifies the expected reward for transitioning from one state to another via a particular action, $r_{t+1} = R(s_t, a_t, s_{t+1})$, and $\gamma$ constitutes a reward discount factor with $\gamma \in [0, 1]$. The POMDP extension involves maintaining a probability distribution over the set of possible states based on observations the agent receives from the environment with every action it takes.

To provide assistance to the human worker while handling state estimation uncertainty, I use a hybrid policy search algorithm that combines techniques well-suited for continuous state and action spaces with value-function based techniques that work well for discrete, lower-dimensional spaces. The scenario of focus is human-robot teaming, where a person is constructing a piece of furniture. The system is provided with a hierarchical task tree [8] encapsulating information about subtasks, primitive actions, and ordering constraints.

The input to the algorithm is data acquired from a motion capture system providing the coordinates of the person's hands in the work area, as well as those of all objects relevant to the task (the different components needed to build the chair). These objects are provided to the system beforehand, together with the hierarchical task tree. The tree is used to learn policies at different levels of the task. For the low-level tree nodes, the policies are learned based on this continuous state space and map the states to particular actions that can be executed within the context of that sub-state. This is included in the first goal of the work (providing supportive behaviors) and focuses on the actual implementation of these actions (e.g. within subtask $a$, help person by stabilizing a piece of wood when needed).

For the high-level tree nodes, policies are learned based on a transformation from the continuous state space provided by the motion capture system data to discrete states representing the state of the human-robot interaction at a high level (e.g. person performing subtask $a$ while robot performing subtask $b$). These policies map the high-level states to actions that represent high-level goals for the robot. Such goals include actions meant to help the human worker with a particular subtask, and actions that direct the robot to work on a different subtask (both of which are included in the first goal of the work), as well as clarification actions meant to disambiguate state information (included in the second goal, that of reducing state estimation uncertainty via communicative actions). Examples of the former are "help with subtask $a_1$," ..., "help with subtask $a_n$," where $n$ is the total number of subtasks, while examples of the latter include directly asking the user questions (e.g. "Are you currently performing subtask $a_1$?").

## IV. Past, Current and Future Work

The initial phase of finding policies that accomplish the two goals mentioned above consisted of investigating how to assess different agents' skill level at performing various tasks. To this end, I was part of a project that investigated how to autonomously predict the amount of time different agents take to complete actions part of common assembly tasks [9]. The project focused on building a model to estimate such durations based on the composition of a skill experience curve (modeling changes in duration due to the agent gaining familiarity with a task via repetition), an agent's estimated tool proficiency, and an agent's estimated motor skill proficiency.

I am currently working on a policy search approach that makes use of this method of predicting an agent's action-completion duration in order to choose appropriate actions corresponding to the robot's high-level goals, as described above. Here, the approach employs a POMDP that maintains a probability distribution over the set of possible states by looking at the observations the robot receives from the environment when it performs an action. When the robot's belief in the true state is too uncertain, it takes information gathering (communicative) actions to improve the current state estimate. When this is not needed, the predicted duration is used to choose what subtask the person might need help with (e.g. durations with high estimated values would need the most amount of help). This component tackles the problem of endowing the robot with the capability of choosing between actions relevant for the two goals presented in this work.

Future work encompasses tackling policy search for low-level nodes of the task tree, looking at how to actually implement actions meant to provide support (i.e. how the robot could actually help with a particular subtask once it knows it needs to). At the low-level, the approach uses an MDP with a continuous state space, based on the input from a motion capture system. The policies resulted from solving MDPs that tackle continuous state and action spaces at the low level and from solving POMDPs that handle discrete state and action spaces at the high level are tied together by the use of the hierarchical structure of the task tree.

## References


[1] M. P. Deisenroth, G. Neumann, J. Peters *et al.*, "A survey on policy search for robotics." *Foundations and Trends in Robotics*, pp. 1–142, 2013.

[2] H. Van Hasselt, "Reinforcement learning in continuous state and action spaces," in *Reinforcement learning*. Springer, 2012, pp. 207–251.

[3] A. G. Barto and S. Mahadevan, "Recent advances in hierarchical reinforcement learning," *Discrete Event Dynamic Systems*, vol. 13, no. 1-2, pp. 41–77, 2003.

[4] J. Pineau, G. Gordon, and S. Thrun, "Policy-contingent abstraction for robust robot control," in *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers Inc., 2002, pp. 477–484.

[5] I. R. Nourbakhsh, K. Sycara, M. Koes, M. Yong, M. Lewis, and S. Burion, "Human-robot teaming for search and rescue," *Pervasive Computing, IEEE*, vol. 4, no. 1, pp. 72–79, 2005.

[6] J. Shah, J. Wiken, B. Williams, and C. Breazeal, "Improved human-robot team performance using chaski, a human-inspired plan execution system," in *International conference on Human-robot interaction, 2011*. ACM, 2011, pp. 29–36.

[7] D. Few, D. J. Bruemmer, Walton *et al.*, "Improved human-robot teaming through facilitated initiative," in *International Symposium on Robot and Human Interactive Communication, 2006*. IEEE, 2006, pp. 171–176.

[8] B. Hayes and B. Scassellati, "Effective robot teammate behaviors for supporting sequential manipulation tasks," in *International Conference on Intelligent Robots and Systems, 2015*. IEEE, 2015, pp. 6374–6380.

[9] B. Hayes, E. C. Grigore, A. Litoiu, A. Ramachandran, and B. Scassellati, "A developmentally inspired transfer learning approach for predicting skill durations," in *International Conference on Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014*. IEEE, 2014, pp. 181–186.