# Training Manned-Unmanned Teams via Curriculum-based Learning

Shiraj Sen, Bradford Miller, Johan Reimann, Brandon Good, Steven Gray, John Hoare and Charles Theurer
*GE Research, {shiraj.sen, millerb, reimann, goodb, steven.gray, john.hoare, theurer}@ge.com*

***Abstract*— Manned-Unmanned teams will be an important component of future factories, hospitals, and battlefields. However, the complexity of these environments and the skills required introduce several challenges and require fundamental change in how we design and build such teams. Presently, many of these teams require humans to teleoperate robots, placing cognitive load on the human operator. Even with well-developed human-robot interfaces, robots working with humans must have certain capabilities to be considered team players, including the ability to recognize intentions of their human teammates and act accordingly. Furthermore, trust is an important component in any successful team and the system design process should allow robots to systematically gain the trust of their human teammates. This paper describes a framework we are building for training human-robot teams where humans can impart their knowledge to robots through guidance and increasingly difficult challenges, and robots can use the interactions to learn the intentions of their human teammates and acquire new skills.**

## I. Overview

There are a number of issues in building effective manned-unmanned teams, including the assembly of behaviors human team members trust to work in relevant conditions and context. A general robot model might include the ability to execute trained behaviors and chain them together to solve problems. However, while we can program a system to perform in a particular domain and problem instance, our goal is to have robots that can demonstrate increasing level of competency over its lifetime with no programming, constructing reusable behaviors that can be pulled from a database, and selecting the right behavior for the problem. To that end, we are investigating a framework (Figure 1) allowing an engineer to synthesize a learning curriculum given a domain description and particular performance concerns. The key idea is that the engineer can specify a series of synthetic worlds for a simulator that will not only allow a learner to develop an effective control policy, but also test the policy's effectiveness against variation through specification of multiple simulation instances. Each simulated trial is recorded and checked, such that if the lesson is not learned the curriculum is updated to add additional learning opportunities, possibly targeted at areas where the learner has difficulties. Additionally, each world can be marked with semantic metadata indicating particular contextual or domain features that may be critical to the policy learned after inspection by the engineer. Furthermore, a human operator conveys a lesson's goal by demonstrating, either physically or telerobotically [1], how to perform the task and have the robot infer the human's intent. The goal is to simplify construction of increasingly sophisticated libraries of behaviors that are competent in not just common scenarios, but rare

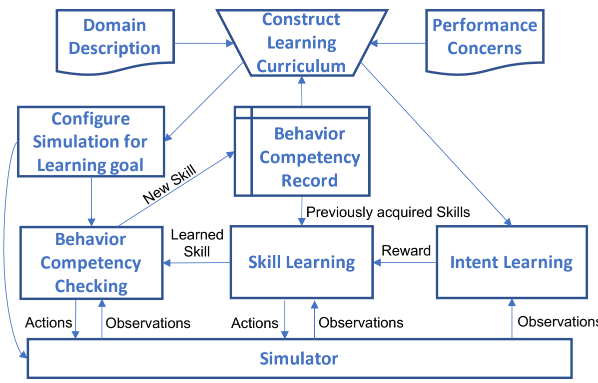


**Fig. 1:** A architecture for using curriculum-based learning to train robots in human-robot teams.

ones as well. In the following sections, we address how we are building the various components of the learning system.

## II. Curriculum-based Learning

When we are in school, we do not start with the most difficult problems for which we have no tools to solve, but with simple problems that allow us to build up an increasingly sophisticated "toolbox" of capabilities that a teacher leads us to apply to harder problems. The idea of having a specific learning curriculum is not a new one, but it is not one that has been widely applied to machine learning [2], however, recent results [3] give us confidence that such an approach should be advantageous both in reducing the time to learn a complex concept and in creating reusable, learned "subconcepts" that can be recombined into more sophisticated concepts. Because we will perform our learning in simulation, it is important to make it easy to specify simulation characteristics as well as predicates around the behavior's competency to be checked – similar to a teacher's lesson plan and a quiz to check progress. After getting a result, additional drills (simulations) can be scheduled or the next step in the learning ladder can be targeted until a final goal has been reached. And unlike children, we can ensure intermediate rungs will be remembered and applied to future situations by recording policies and metadata describing the conditions under which they were created.

### *A. Leader/Follower Scenario*

Assume two robots: A and B. Our task is to train B to follow A, regardless of A's behavior. First we specify our ultimate goal, that B follows A in varying terrain with denied communications and limited visibility. The teacher now might divide the problem into various kinds of represented complexity, e.g., path vs. rough terrain, communications

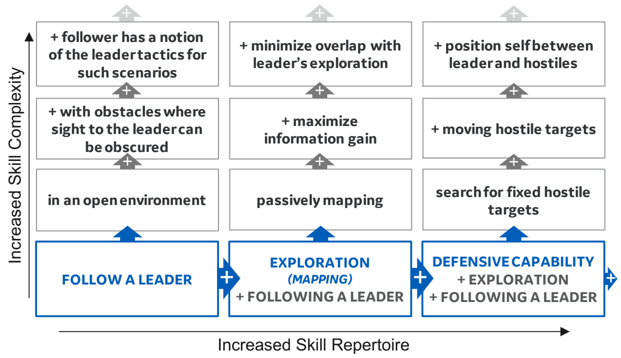


**Fig. 2:** A hierarchy of skills as part of curriculum-based learning.

restrictions, and visibility. The simplest test might be that B can follow A's path in an empty world with its sensors, and criteria for progressing to the next lesson is a successful run showing B tracking A for the entire duration. Next steps might be to add following at a distance, then splitting the path A takes so B has to replicate A's path choice, etc. Eventually we begin adding visual obstacles that prevent B from always being able to observe A. B would have to make predictions about A's behavior when not directly observed, presumably from prior observed behavior. Each time we must specify the kind of world we want to simulate, add in specific (though possibly randomized) behavior for robot A to ensure the lesson is conveyed, and the aspects of B's behavior to be checked for success. Note that this need not be identical to B's scoring function. For instance our test might check if B is "fooled" by a feint A makes while it is observed by B before selecting a path when obscured. In this case failure might indicate to the teacher that the curriculum needs to be adjusted to include more deceptive acts from A, but not that B's scoring function is faulty. In other words, some tests of B's competency are really about B's internal model, while others might be indicative of an incorrectly learned prior lesson. We can see similar kinds of misplaced confidence in [4].

Figure 2 describes the skills being considered as part of the curriculum. We apply our learning framework to the problem of following a leader in an urban environment. In this curriculum, each column describes a new skill that reuses a previously acquired skill from a previous column. Skill complexity increases by row, making the environment or task objective more complicated. We created an environment in Open AI Gym [5] that interacts with a Gazebo-based simulator using ROS. This interface allows a robotic agent to be controlled and trained using machine learning.

Our current effort in this area is two-fold: 1) How to specify the policy learned by the learner in a reusable way. We are experimenting with extensions of a "Visual Cognitive Computer" metaphor [3] to support human advice, such as described in [6]. 2) More importantly, we are looking toward a higher level specification language to establish the learning goals, tests, and simulation configuration that would make the task more accessible. Here we are inspired by the more basic concepts of describing and understanding the world around us, such as [7], to move away from programming currently required.

## III. Skill and Intent Learning

A central challenge in building human-robot teams is enabling the robotic team members to quickly handle new situations. Having a robot repeatedly make the same mistake in an unexpected situation not only results in inefficient task execution, but also frustrates human collaborators. Thus, it is critical to endow robotic team members with the ability to learn new skills or to revise old ones. One avenue for exploration of new skills is to frame the problem as a reinforcement learning problem. Given a desired outcome expressed through a reward function, the robotic team members will search for an optimal policy that prescribes actions to take given the current state and previous actions to accomplish a goal. The advantage of relying on reinforcement learning to learn new skills, as opposed to recording an optimal policy from an expert and playing it back, is that the robots continue to search for better ways to accomplish tasks and learn how to behave in novel situations where expert input is not available. However, several key research challenges must be addressed to ensure skills are learned effectively:

- Robots must learn from few examples to avoid unnecessary operational inefficiencies [8].
- The learning process must be safe; the robots can adapt their behavior without incurring damage [9].
- If using simulation learn new skills, those skills must transfer to real environments [8].
- Even though a learned policy may solve the task, it may not be in a desirable way to a human observer or teammate [10], reducing overall team effectiveness.

The need for "reward engineering" [11] or how to capture a desired outcome in a reward function is an additional challenge. To address this problem, the robots will have the ability to learn behaviors from a human operator through an inverse reinforcement learning (IRL) algorithm. IRL is designed to solve the following problem: Given several demonstrations from a teacher, the state and action examples are captured and used to recover/discover a reward function. In contrast to behavioral cloning, where a skill or behavior is learned directly from a user in a supervised setting, the inverse reinforcement learning approach captures high-level goals. The compact representation of these goals allows the robot to experiment with different policies to accomplish the task, thereby allowing the robots to potentially correct for suboptimal policies provided by the teacher.

In high-performing manned-unmanned teams, the robot(s) must be able to make sense of the actions and intent of the teammate(s) [10]. Encapsulated in the increased skill complexity tasks of Figure 2 are natural and effective reactions to the changing environment, teaming arrangements, and changing team goals. By having a mixture of subject matter experts and end users participate in the end-to-end teaching of the robotic team mates, the ability to learn the correct skills relevant to the mission, activities, and contexts can be achieved.

# REFERENCES


[1] J. R. Hoare, J. Foehner, S. R. Gray, S. Sen, H. Tan, and K. Macdonald, "Telerobotic control of autonomous mobile robots for industrial applications using virtual reality," in *RO-MAN Workshop on Human-Robot Interaction through Virtual Reality (HRI-VR)*, 2018.

[2] J. Beal, R. Paul, and L. Robert, "Curricula and metrics to investigate human-like learning," in *AAAI Spring Symposium: Agents that Learn from Human Teacher*, 2009, pp. 16–19.

[3] M. Lázaro-Gredilla, D. Lin, J. S. Guntupalli, and D. George, "Beyond imitation: Zero-shot task transfer on robots by learning concepts as cognitive programs," *CoRR*, vol. abs/1812.02788, 2018. [Online]. Available: http://arxiv.org/abs/1812.02788

[4] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2015.

[5] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," *CoRR*, vol. abs/1606.01540, 2016.

[6] N. Mavridis and D. K. Roy, "Grounded situation models for robots: Bridging language, perception, and action," in *AAAI-05 workshop on modular construction of human-like intelligence*, 2005.

[7] J. M. Mandler, "How to build a baby: ii. conceptual primitives," *Psychological review*, vol. 99, no. 4, 1992.

[8] H. Mania, A. Guy, and B. Recht, "Simple random search of static linear policies is competitive for reinforcement learning," in *Advances in Neural Information Processing Systems 31*, 2018, pp. 1800–1809.

[9] N. Fulton and A. Platzer, "Safe reinforcement learning via formal methods: Toward safe control through proof and learning," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018, pp. 6485–6492.

[10] J. R. Hoare and L. E. Parker, "Using on-line conditional random fields to determine human intent for peer-to-peer human robot teaming," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 4914–4921.

[11] D. Dewey, "Reinforcement learning and the reward engineering principle," in *AAAI Spring Symposia*, 2014.