Online Learning in Repeated Human-Robot Interactions

Vahan Babushkin, Mayada Oudah, Tennom Chenlinangjia, Ahmed Alshaer, Jacob W. Crandall
Masdar Institute of Science and Technology
Abu Dhabi, UAE
jcrandall@masdar.ac.ae

Abstract

Adaptation is a critical component of collaboration. Nevertheless, online learning is not yet used in most successful human-robot interactions, especially when the human's and robot's goals are not fully aligned. There are at least two barriers to the successful application of online learning in HRI. First, typical machine-learning algorithms do not learn at time scales that support effective interactions with people. Algorithms that learn at sufficiently fast time scales often produce myopic strategies that do not lead to good long-term collaborations. Second, random exploration, a core component of most online-learning algorithms, can be problematic for developing collaborative relationships with a human partner. We anticipate that a new genre of online-learning algorithms can overcome these two barriers when paired with (cheap-talk) communication. In this paper, we overview our efforts in these two areas to produce a situation-independent, learning system that quickly learns to collaborate with a human partner.

1 Introduction

In ad hoc teams (Stone et al. 2010), assistive robotics, manufacturing systems, and other collaborative domains, robots must establish long-term collaborative relationships with human partners. Interactions in these domains often have three characteristics. First, as in repeated interactions between people, these human-robot interactions are often punctuated by conflicting interests – the human's and robot's goals are not always fully aligned. Second, the human's behaviors and tendencies are initially unknown to the robot and to designers of the robotic system. Third, people typically adapt as they interact with the robot (HRI Community Page 2012). These three characteristics imply that, in many scenarios, robots must employ online learning to establishing effective collaborations with human partners.

Despite its importance, online learning is a significant challenge to HRI for at least two reasons. First, to date, most machine-learning algorithms designed for generic multi-agent scenarios do not learn at time scales that support effective interactions with people, especially when the human partner is also adapting to the robot. Algorithms that do learn

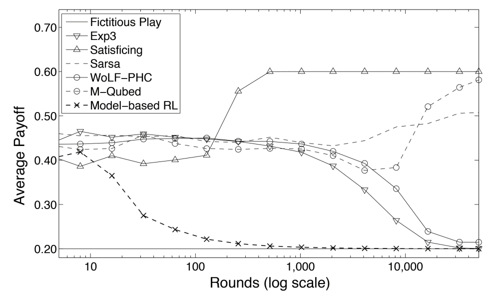
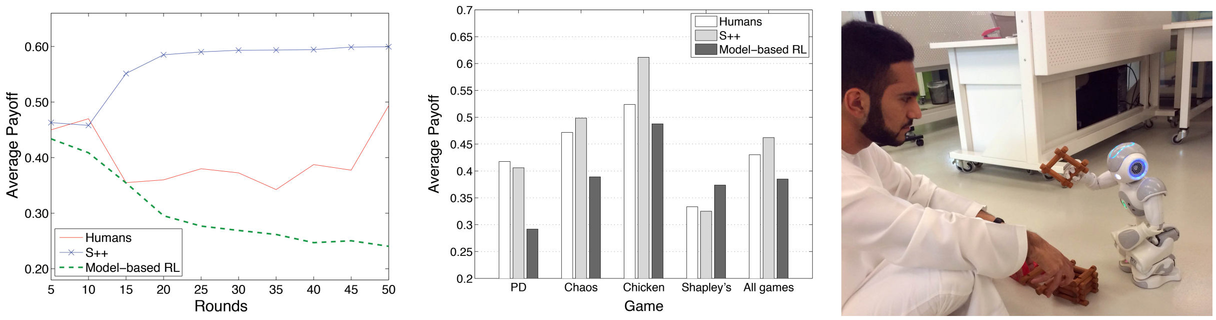


Figure 1: Average performance of existing online-learning algorithms in self play in a repeated prisoners' dilemma. A payoff of 0.60 results from mutual cooperation, 0.20 from mutual defection. Most online-learning algorithms either fail to learn the cooperative solution, or take hundreds to thousands of rounds of experience to learn a collaborative solution (e.g., Satisficing and M-Qubed).

at appropriate time scales for human interaction are usually either customized for specific scenarios, or learn myopic strategies that do not lead to effective long-term collaborations. These statements are true for even very simple scenarios, such as the repeated prisoners' dilemma (Figure 1). Second, random exploration, a core component of traditional learning algorithms, can be problematic for the development of a collaborative relationship with a human partner.

Our proposed solution for overcoming these barriers is two-pronged. First, we are developing a new family of online-learning algorithms for repeated stochastic games. Our intent is to develop algorithms that learn at much faster time scales. In the absence of explicit communication, we have shown that these algorithms quickly learn effective collaborative solutions when associating with other online-learning algorithms in repeated matrix games (Crandall 2014). Via user study (Ishowo-Oloko et al. 2014), we have also shown that the performance of these algorithms rival human capabilities in repeated games of relatively brief durations (Figures 2a-b). We have recently extended these algorithms to repeated stochastic games. Evaluations are currently underway to assess the effectiveness of these algorithms to interact with people in the absence of explicit communication.

However, when people interact with each other and with robots, they utilize cheap talk and other forms of commu-



(a) In self play in a prisoner's dilemma (b) Associating with humans (c) Example collaborative scenario

Figure 2: (a-b) Average payoffs obtained in a user study pairing online-learning algorithms and people; see Ishowo-Oloko (2014) for details. (c) A collaborative human-robot scenario in which we are evaluating our online-learning algorithms.

nication to help facilitate collaborative behavior. Thus, the second component of our work is to integrate a scenario-independent communication method with online learning. We anticipate that this will (1) allow a robot to coordinate collaborative behavior with others at an even faster rate and (2) mitigate the negative effects of exploratory actions.

We briefly overview our work in these two areas.

2 Learning to Collaborate with People

Machine-learning algorithms have been used in several forms of human-robot interaction. For example, machine learning is a core component of learning from demonstrations (Thomaz and Breazeal 2008; Argall et al. 2009). However, in such interactions, the robot's goals arise from the goals of the human with which it interacts – there is no conflicting interest. Furthermore, learning from demonstrations is typically used as a method for learning a skill (offline), which will then later be used by the robot. It is not often used during real-time human-robot collaborations.

Online learning has been used successfully in scenario-specific human-robot interactions. For example, in the works of Hoffman and Breazeal (2008) and Nikolaidis et al. (2014), online learning was used to build a model of the human partner's behavior. Our work differs from these works in that (1) we seek situation-independent online-learning algorithms and (2) we anticipate that the robot and the human may not have identical preferences. Thus, our problem of interest is closely aligned with the challenge of multi-agent learning (Shoham, Powers, and Grenager 2007), but where one of the players is a human. The goal of both the human and the robot is to find a strategy that maximizes its own individual payoffs (often, but not always, achieved when the players "cooperate").

Since most existing multi-agent learning algorithms fail to learn collaborative strategies within time scales that support interaction with a human partner (Figure 1), we are developing a new family of multi-agent learning algorithms (Crandall 2014). These algorithms are novel expert algorithms that operate on set of experts that implement strategies that represent a variety of ideals, such as the best response, pareto optimality, security, etc. In each round of interaction, the expert algorithm chooses to follow the strategy of one of these experts. Over time, it seeks to learn to always follow the most successful expert for the given scenario.

As mentioned previously, preliminary results indicate that these algorithms are at least as effective as people at establishing collaborative relationships with people in repeated games (Figures 2a-b). In our ongoing work, we have extended these algorithms for repeated stochastic games, which can model rich environments such as those required in human-robot collaborative tasks.

3 Adding Cheap Talk

To date, we have considered online learning in environments that do not permit explicit communication between interactants. However, in collaborative tasks, communication (non-verbal signals and cheap talk) is often possible and useful. Such communication allows humans to more quickly coordinate their behaviors with each other. Thus, determining how to generate and integrate cheap-talk communication with online learning appears to be critical to learning effectively in human-robot collaborations.

Most machine-learning algorithms for stochastic games employ statistical processes that are a "black box" to people. However, the algorithms we consider can show more clarity. Each of the experts used by the expert algorithms computes a particular high-level ideal. These ideals form a high-level behavioral agenda that is understandable to people, and which can be communicated verbally in a situation-independent way. Thus, as the expert algorithm switches between experts, the robot can explicitly communicate its future behavior and its expectations for its collaborator.

4 Embodiment

Our preliminary assessments indicate that such communication will vastly enhance a robot's ability to learn to collaborate with human partners. We are now in the process of evaluating our integrated system via formal user studies. Specifically, we studying interactions between people and a Nao robot in several collaborative tasks. In one of these tasks, the robot and human must learn to collaboratively construct a (toy) building structure (Figure 2c). Other scenarios involve determining how to share limited resources, such as blocks and space.

References

Argall, B. D.; Chernova, S.; Veloso, M.; and Browning, B. 2009. A survey of robot learning from demonstration. *Robotics and Autonomous Systems* 57:469–483.

Crandall, J. W. 2014. Towards minimizing disappointment in repeated games. *Journal of Artificial Intelligence Research* 49:111–142.

Hoffman, G., and Breazeal, C. 2008. Achieving fluency through perceptual-symbol practice in human-robot collaboration. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction*, 1–8.

HRI Community Page. 2012. Adaptation, learning, and training. http://humanrobotinteraction.org/adaptation-learning-and-training/. Accessed: 2014-07-13.

Ishowo-Oloko, F.; Crandall, J. W.; Cebrian, M.; Abdallah, S.; and Rahwan, I. 2014. Learning in repeated games: Human versus machine. *arXiv:1404.4985*.

Nikolaidis, S.; Gu, K.; Ramakrishnan, R.; and Shah, J. 2014. Efficient model learning for human-robot collaborative tasks. *arXiv:1405.6341v1*.

Shoham, Y.; Powers, R.; and Grenager, T. 2007. If multi-agent learning is the answer, what is the question? *Artificial Intelligence* 171.

Stone, P.; Kaminka, G. A.; Kraus, S.; and Rosenschein, J. S. 2010. Ad hoc autonomous agent teams: Collaboration without precoordination. In *Proceedings of the Twenty-Fourth Conference on Artificial Intelligence*, 1504–1509.

Thomaz, A. L., and Breazeal, C. 2008. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence* 172:716–737.