Leveraging Human Inputs in Interactive Machine Learning for Human Robot Interaction

Emmanuel Senft
Centre for Robotics and Neural Systems
Plymouth University, UK
emmanuel.senft@plymouth.ac.uk

Séverin Lemaignan
Centre for Robotics and Neural Systems
Plymouth University, UK
severin.lemaignan@plymouth.ac.uk

Paul E. Baxter
Lincoln Centre for Autonomous Systems
University of Lincoln, UK
pbaxter@lincoln.ac.uk

Tony Belpaeme
CRNS – Plymouth University, UK
imec - ID Labs – Ghent University, Belgium
tony.belpaeme@plymouth.ac.uk

ABSTRACT

A key challenge of HRI is allowing robots to be adaptable, especially as robots are expected to penetrate society at large and to interact in unexpected environments with non-technical users. One way of providing this adaptability is to use Interactive Machine Learning, i.e. having a human supervisor included in the learning process who can steer the action selection and the learning in the desired direction. We ran a study exploring how people use numeric rewards to evaluate a robot's behaviour and guide its learning. From the results we derive a number of challenges when designing learning robots: what kind of input should the human provide? How should the robot communicate its state or its intention? And how can the teaching process by made easier for human supervisors?

Keywords

Interactive Machine Learning; Autonomy; HRI

1. INTRODUCTION

One important challenges in HRI is to allow users, who often have no technical expertise, to personalise the behaviour of the robot they are using. It seems infeasible to expect either users to be satisfied by a robot with a static behaviour or for the robot's designers to be able to anticipate all the needs of the users and all the different environments a robot could interact in. For this reason, we argue that robot behaviour should be adaptive at run-time, and especially that non-experts in technology should be able to teach a robot new action policies.

Interactive Machine Learning (IML) is a field of research which aims to include end-users in the machine learning process [1, 3]. The idea is to move away from robots as
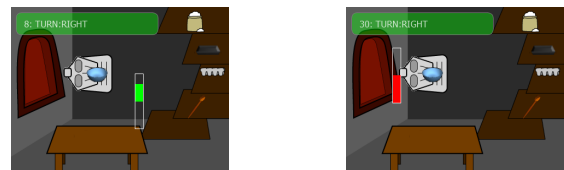
Figure 1: Examples of positive (left) and negative (right) reward in the robot cake baking task.

complex black boxes with inaccessible input and output, to systems that can be intuitively (re)programmed by the users. One advantage of this approach is it empowers users with the ability to personalise their robot according to their needs and desires.

IML has principally been tested on virtual agents. A good example of IML is the TAMER framework [4] which predicts the reward a human would give and use this prediction to select a next action maximising the predicted reward.

2. EXPERIMENT

Methodology.

IML has not often been applied to robotics. An application using virtual robots was presented in [6], which presented a study where we compared two different methods used to teach robots an action policy. The first, Interactive Reinforcement Learning (IRL), is derived from Reinforcement Learning (RL) [7], the difference being that user now provides rewards, rather than the environment (cf. Figure 1). In our implementation, the participant could evaluate the robot's actionsby moving a slider on a graphical interface, the value of the slider acted as the reinforcement learning reward. The second method, inspired by [5], uses a more direct control method in which the robot communicates its intentions to the participant who can either passively accept the suggestion or actively select an alternative action.

In the task, inspired by Thomaz and Breazeal [8], a virtual robot is in a kitchen and has to learn how to bake a cake. The users know what the robot should do to finish the cake, but multiple strategies can lead to success. The IRL algorithm is similar to that used by [8].
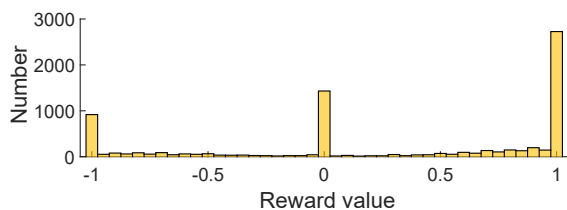
Figure 2: Distribution of rewards according to their numeric value.

Participants were divided into two groups of 20, each group interacting with a different system.

Results.

This paper reports how participants used the rewards in the IRL condition. Figure 2 presents the distribution of rewards given to the robot. In our case, the distribution was tri-modal, with only three types of rewards given to the robot. Over a total of 7364 rewards, we observe 899 with a value of -1, 1337 with a value of 0 and 2653 of a value of 1.

This indicates that even if participants had the opportunity to provide fine grained numerical rewards, they decided to evaluate the robot's actions either as *bad*, *neutral* or *good*. We identify three reasons potentially explaining this behaviour. The first one is related to the unambiguity of the expected behaviour: when a desired strategy is clear, humans might only use extremes to give feedback to the robot. Alternatively, the tri-modality of the reward distribution can be due to the interface used, a slider makes the use of extremes easier. And lastly, as participants were time constrained (they only had 2 seconds to evaluate the robot's action) they might not have taken the time to use a fine grained rewarding strategy.

3. DISCUSSION

From these results, we derive three challenges that robot designers will face when allowing humans to teach robots.

Type of inputs.

The first challenge is to make human input efficient and generalisable over different tasks. RL seems like a reasonable approach: the user can provide numerical rewards to evaluate the action executed by the robot. However, as the task becomes more complex, the algorithm converges only after a long series of trials and errors which is undesirable. Another limit of numerical rewards is that they generally are assigned *after* the execution of an action, and so do not allow the supervisor to prevent the robot from making an error, even if the supervisor could have known beforehand that this action was not appropriate. Reward-based learning is general, but it does not make good use of human domain knowledge and tutoring competency.

Other types of inputs could be used; in [6], we propose using commands rather than feedback. Commands allow the user to have more control of the robot, but limit the actions to a predefined set of actions. A way to generalise commands to a larger set would be to use natural language and ability to teach new actions associated to new commands. A robot could also combine different types of inputs from the human: both explicit (rewards or commands for example) and implicit (such as the reactions of other humans).

Clarity of robot's communication.

The robot should also provide the human with feedback about its internal state, including its intentions, uncertainty, learning progress and confidence. In [6], we argue that intention communication is especially important when robots are interacting in the real world, so as not to fluxom people or execute undesired actions. Furthermore, if the robot has planning abilities, the robot can also explain its actions and communicate what its next actions will be.

Similarly to the content of the robot's communication, the medium is important. Humans have evolved to use social signals, and robots should use these too. Speech could be a good way of communicating more complex states, intentions and plans, but it would be interesting to modulate the sentences expressed by the robot in a way which is socially acceptable and which does not annoy the long-term user, avoiding repetitions and bluntness often associated with robots.

Reduce the workload on human teachers.

A last important challenge is maintaining the user's comfort when teaching. As explained in [2], a robot using the same mode of communication without considering the human it is interacting with could annoy the user. In our study, many participants also reported frustration due to them often knowing what the robot should do, but not being able to have the robot execute the desired action. Robots learning from humans should reduce the workload on the teacher, and give the teacher enough control of the robot's actions while taking into account the human's state when learning.

Acknowledgments

4. REFERENCES

[1] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza. Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4):105–120, 2014.

[2] M. Cakmak and A. L. Thomaz. Designing robot learners that ask good questions. In *HRI'12*, pages 17–24, 2012.

[3] J. A. Fails and D. R. Olsen Jr. Interactive machine learning. In *IUI'03*, pages 39–45, 2003.

[4] W. B. Knox and P. Stone. Interactively shaping agents via human reinforcement: The tamer framework. In *K-CAP'09*, pages 9–16, 2009.

[5] E. Senft, P. Baxter, J. Kennedy, and T. Belpaeme. SPARC: Supervised progressively autonomous robot competencies. In *ICSR'15*, pages 603–612, 2015.

[6] E. Senft, S. Lemaignan, P. E. Baxter, and T. Belpaeme. Sparc: an efficient way to combine reinforcement learning and supervised autonomy. In *FILM Workshop at NIPS'16*, 2016.

[7] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

[8] A. L. Thomaz and C. Breazeal. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence*, 172(6):716–737, 2008.