Object-Focused Advice in Reinforcement Learning (Extended Abstract)

Samantha Krening
Georgia Institute of Technology
skrening@gatech.edu

Brent Harrison
Georgia Institute of Technology
brent.harrison@cc.gatech.edu

Karen M. Feigh
Georgia Institute of Technology
karen.feigh@gatech.edu

Charles Isbell
Georgia Institute of Technology
isbell@cc.gatech.edu

Andrea Thomaz
Georgia Institute of Technology
athomaz@ece.utexas.edu

ABSTRACT

In order for robots and intelligent agents to interact with and learn from people with no machine-learning expertise, robots should be able to learn from natural human instruction. Many human explanations consist of simple sentences without state information, yet most machine learning techniques that incorporate human guidance cannot use non-specific explanations. This work aims to learn policies from a few sentences that aren't state specific. The proposed Object-focused advice links an object to an action, and allows a person to generalize over an object's state space. To evaluate this technique, agents were trained using Object-focused advice collected from participants in an experiment in the Mario Bros. domain. The results show that Object-focused advice performs better than when no advice is given, the agent can learn where to apply the advice in the state space, and the agent can recover from adversarial advice. Also, including warnings of what not do to in addition to advice of what actions to take improves performance.

CCS Concepts

•Human-centered computing → Text input; •Computing methodologies → *Reinforcement learning;*

Keywords

Advice; Reinforcement Learning; Human Teachers

1. INTRODUCTION

This work focuses on an area of learning from explanations that has been addressed little in previous research - how to learn from human explanations that lack state information. Learning from a few simple sentences is a worthwhile goal because it can decrease the amount of time and effort a human teacher needs to contribute compared to demonstrations or critique. Most forms of machine learning that use human input force the person to provide state-specific information, whether or not that level of detail is reasonable or intuitive, and cannot learn from human sentences that lack state information [1, 4, 2, 6, 5].

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*
Copyright © 2016, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

2. OBJECT-FOCUSED HUMAN ADVICE

This work proposes **Object-focused advice**, a method in which human advice is tied to objects instead of specific states and is generalized over the object's state space. Consider this explanation from the popular Super Mario Brothers game: "Mario should jump on enemies." While this advice would easily be understood by a human student, it proves problematic for reinforcement learning agents. The teacher did not specify state information like where the enemy needs to be with respect to Mario and what Mario's velocity should be. Knowing that Mario should jump on an enemy is valuable information, but how can an agent make use of it if no state information is provided?

Object-focused advice addresses this by linking an *object* to an action that should be used around that object. We define this to be **advice** because it tells the agent what actions to take.

> Mario should jump to collect *coins*, and jump over *chasms*.

A person might also provide **warnings** in an explanation to teach the agent what actions to avoid.

> Do not move right into an *enemy*. Do not fall into *chasms*.

Using Object-focused advice that is independent of the object's state allows the person to perform object-level generalization instead of the agent. Generalizing over the entire state space of an object may seem drastic, but it is a way to quickly operationalize human explanations without state information. It is unrealistic to expect people to provide detailed state information when giving advice. A person might say, "Jump on the enemy," but will not say, "Hold the jump key for 10 frames when Mario is within 2.5 horizontal blocks of an enemy with a velocity of 3.2 units/frame." The agent will take the action advice of "Jump on the enemy," and determine to which portions of the state space, if any, the advice applies.

Following advice a set number of times and then relying on experience allows the agent to recover from adversarial

advice, which is input that is expected to result in minimal reward. Also, Object-focused advice lets the agent's 'trust' in the human advice vary across the domain. The agent will treat each piece of advice without prejudice; if a person provides one piece of good advice along with eight pieces of bad advice, the agent will use its experience to build policies that reflect the good and ignore the bad.

Advice describes what to do, while warnings describe what not to do. To incorporate warnings, all objects in the state space are taken into account together by summing up the Q-values associated with each action across all objects. Choosing an action by taking multiple objects into account allows us to get an idea of the overall severity of each action.

3. RESULTS

The experiment was conducted in the Mario Bros. domain [7] It is a partially-observable environment in which Mario must collect rewards and avoid being harmed while moving toward the goal. During the experiment, we collected responses from participants in which one action was advised for every object; this advice was used to train agents using Object-focused advice and Object-focused Q-learning [3].

Figure 1 shows an agent trained with adversarial advice quickly recovers and performs as well as no advice, but not as well as good advice. After 400 trials, the best advice from the experiment led to Mario falling into chasms approximately 16% of the time, while the agents using adversarial or no advice fell into chasms 34% of occurrences. The results were averaged over 100 episodes with a sliding window average of width of 25. The parameters used were $\alpha = 0.1$, $\gamma = 0.95$, $\epsilon_0 = 0.8$, and $\epsilon_{min} = 0.15$. ϵ-greedy exploration was used.

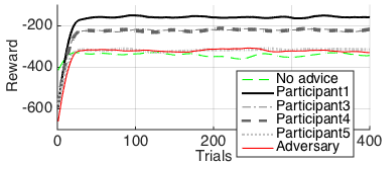

Figure 1: Cumulative Reward for Chasms.

Figure 2 shows the agent learns to which part of the state space the advice applies. The agent was advised to jump to the right when encountering Goombas. The agent learned this was a good policy when the Goomba was to the right of Mario in a 'goldilocks' zone - not too close but not too far away - but was bad advice when the enemy was above Mario.

Incorporating "what not to do" warnings in addition to "what to do" advice approximately doubled the cumulative reward earned by the agent for different objects.

4. CONCLUSIONS

We presented a novel form of human advice and warnings that links objects to actions. Advice and warnings are object-specific, and so do not require people to specify state variables. Using warnings in addition to advice and combining the Q-values from multiple objects improves the agent's performance. Object-focused advice allows people to generalize over an object's state space, which lets agents learn from a few simple sentences with no numbers or particulars describing the state. A model-free approach has been described that does not require the intensive construction of formal language translations.

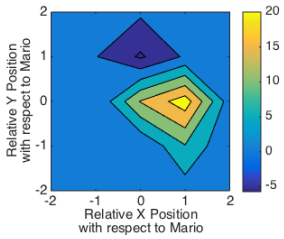

Figure 2: Visualizing Object-level Generalization in a Policy for Goombas. The color scale represents Q-values showing when to jump right.

The goal of Object-focused advice is not to capture all the nuances and subtleties of free-form teaching, but rather to make use of human explanations without state information. It is vital to develop methods that use human explanations that aren't state-specific since they reflect much of non-expert instruction.

5. ACKNOWLEDGMENTS

This work was funded under ONR grant number N000141410003.

REFERENCES

[1] B. D. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483, 2009.

[2] S. Chernova and A. L. Thomaz. Robot learning from human teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 8(3):1–121, 2014.

[3] L. C. Cobo, C. L. Isbell, and A. L. Thomaz. Object focused q-learning for autonomous agents. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 1061–1068. International Foundation for Autonomous Agents and Multiagent Systems, 2013.

[4] S. Griffith, K. Subramanian, J. Scholz, C. Isbell, and A. L. Thomaz. Policy shaping: Integrating human feedback with reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 2625–2633, 2013.

[5] J. MacGlashan, M. Babes-Vroman, M. desJardins, M. Littman, S. Muresan, S. Squire, S. Tellex, D. Arumugam, and L. Yang. Grounding english commands to reward functions. In *Proceedings of Robotics: Science and Systems*, Rome, Italy, July 2015.

[6] R. Maclin, J. Shavlik, L. Torrey, T. Walker, and E. Wild. Giving advice about preferred actions to reinforcement learners via knowledge-based kernel regression. In *Proceedings of the National Conference on Artificial intelligence*, volume 20, page 819. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.

[7] J. Togelius, S. Karakovskiy, and R. Baumgarten. The 2009 mario ai competition. In *Evolutionary Computation (CEC), 2010 IEEE Congress on*, pages 1–8. IEEE, 2010.