Teaching Social Behavior through Human Reinforcement for Ad hoc Teamwork - The STAR Framework

Extended Abstract

Shani Alkoby
The University of Texas at Austin
Austin, Texas, USA
shani@cs.utexas.edu

Avilash Rath
The University of Texas at Austin
Austin, Texas, USA
rathavilash@gmail.com

Peter Stone
The University of Texas at Austin
Austin, Texas, USA
pstone@cs.utexas.edu

ABSTRACT

As AI technology continues to develop, more and more agents will become capable of long term autonomy alongside people. Thus, a recent line of research has studied the problem of teaching autonomous agents the concept of ethics and human social norms. Most existing work considers the case of an individual agent attempting to learn a predefined set of rules. In reality however, social norms are not always pre-defined and are very difficult to represent algorithmically. Moreover, the basic idea behind the social norms concept is ensuring that one's actions do not negatively influence others' utilities, which is inherently a multiagent concept. Thus, here we investigate a way to teach agents, as a *team*, how to act according to human social norms. In this research, we introduce the STAR framework used to teach an *ad hoc team* of agents to act in accordance with human social norms. Using a hybrid team (agents and people), when taking an action considered to be socially unacceptable, the agents receive negative feedback from the human teammate(s) who has(have) an awareness of the team's norms. We view STAR as an important step towards teaching agents to act more consistently with respect to human morality.

KEYWORDS

Ad hoc; Reinforcement Learning; Social Norms

1 INTRODUCTION

Embedding ethics and social norms into AI systems in general, and in the decision making process of autonomous agents in particular, has long been a grand challenge for AI [20]. As a result, many studies focusing on technical approaches for enabling AI systems and autonomous agents to capture the concept of social norms have emerged [22]. However, despite the abundance of research dealing with the question of how to inject social norms into intelligent agents, there are still three main problems that remain unsolved [17]. First, most studies address the problem of a single autonomous AI agent working in isolation. In reality, however, AI agents will increasingly work together in teams that will include humans and other agents as their team members. This discrepancy can lead to many uncertainties and incompatibilities due to policies being tailored to a single agent that may not necessarily be optimal or even relevant in the context of a team. Second, none of the existing studies has offered a way to address the various cultural and temporal dynamics of the broad spectrum of human norms, i.e., the fact that ethics and social norms are not absolute, timeless, universally agreed upon concepts. Third, most approaches use the same formalism for both the function to be maximized and the social boundaries. We note that, even though this can help make the technical calculations easier, due to them being two independent objectives, it may be important to allow for the possibility of goals and norms being represented differently.

In this research we present an approach for teaching autonomous agents the concept of human social norms which provides a solution to the above problems. For this purpose we turn to the "ad hoc teamwork" setting in which a team of agents is formed ad hoc, for a particular purpose, and thus the team strategies cannot be developed a priori. Ad hoc teamwork has been studied recently in the AI literature [2, 7]. However, to date, no attention has been dedicated to examining whether the methods proposed are *safe* in the sense of preventing the agents from choosing socially unacceptable actions in order to complete their task. We use this scenario in which agents are working together and their actions have mutual influence on one another, to create an online mutual learning process which leads to socially acceptable behavior by the entire team.

In this research we introduce a novel training paradigm called "**S**ocially **T**raining **A**gents via **R**einforcement" (**STAR**). Using STAR we study the case of a hybrid team including agents and people. During the cooperation, when taking an action considered to be unacceptable (as opposed to ineffective), the agents receive negative feedback on a dedicated channel for this purpose from the human teammate(s). This will allow online learning of social codes based on the specific cultural and temporal dynamics relevant to the society the agents are part of. Our method builds upon past work introducing the TAMER framework for learning from positive and negative human feedback [15]. TAMER is based on the assumption that feedback is given to teach the agent how to be more effective. Our work differs by introducing a separate channel by which a person can indicate actions that are unacceptable regardless to how effective they are. Using this social feedback during the learning process, agents are able to develop a set of internal rules such that given a task they will be able to solve it compatibly with the humans' concept of social norms.

2 ETHICS AND AI

Humans often constrain their decisions according to some exogenous priorities such as morality, ethics, or religion [19]. Intelligent agents should be able to do the same. Thus, the AI community is interested in building such smart systems that will be able to

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

restrict their actions by similar principles [4, 20, 22]. For example, the work of Balakrishnan et al. [5, 6] studies the problem of applying dynamic ethics rules to content recommendation systems. Another example is the value alignment problem [12–14, 18]. We note that most existing approaches to the value alignment problem assume that misalignment comes from an error in goal specification, inadequate constraints on actions, or lack of human knowledge, whereas we assume the goal specification from the human to be precise and add an additional layer of social norms to the agent's learned behavior. Moreover, most existing work trying to incorporate ethics and social norms into AI agents only considers the case of an individual agent. Those that do consider collective decision making [11], provide only an initial approach to embedding ethical and social codes into collective decision making.

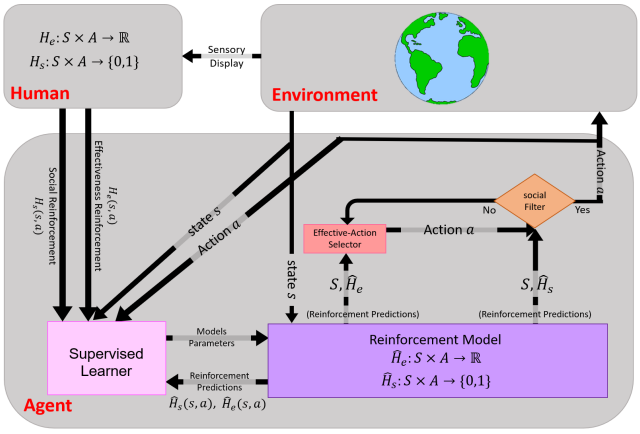

Figure 1: The STAR framework.

3 AD HOC TEAMWORK

The design of autonomous agents that can be a part of an ad hoc team is an important open problem in multiagent systems and as such has been widely studied [7, 16, 21]. Several works addressed this problem by proposing methods which utilize beliefs over a set of hypothetical behaviors for the other agents [1, 3, 8, 9]. One crucial issue that is yet to be studied is the question of *how social the agents' actions are on their way to the goal.* This issue is particularly important due to the fact that we are heading toward a future in which agents will be capable of long-term autonomy without direct supervision of humans. Consider for example, future robots which, as part of their daily tasks, may need to wait in lines with people. In this case, knowing the social norms within the community with regards to queuing behaviors could strongly influence, for example, whether a robot should crowd to the front or wait patiently in line.

4 SOCIAL BEHAVIOR IN AD HOC TEAMWORK

The goal of ad hoc teamwork is for an *individual* agent to figure out how best to act in order to contribute to its team's success, *given the behaviors and/or learning strategies of its teammates.* However, to date, team success has not included any notion of social norms. Thus an ad hoc teammate may take actions that, while in the long-term interest of the team, violate the constraints of social behavior among agents and/or with human teammates. Understanding how to inject social norms into the decision making process of the agents is a timely challenge.

We note that the concept of social norms is a notoriously difficult capability to represent algorithmically [10]. We, therefore, propose that it ought to be taught directly by instructors. Since our objective is to align an agent with subjective human social norms (i.e., we do not rely on there being absolute, universally acceptable norms), humans themselves have the knowledge that can enable the learning process, reducing costly sample complexity.

In this research, we address the problem of controlling the agent's social behavior using the framework of ad hoc teamwork and a novel extension of TAMER which we introduce in the following section.

4.1 The STAR Framework

Like TAMER, STAR uses human feedback. STAR however, does not limit the use of human feedback only to the effectiveness aspect of the action performed, i.e., it has an additional channel by which a person can indicate actions that are unacceptable even when they are technically effective. Based on both the effectiveness signal and the social signal, agents need to find ways to solve the problem that do not violate the social customs. In effect, agents must create a form of "inner conscience" helping them to solve a given problem compatibly with humans' social norms.

Figure 1 shows the interaction between a human, the environment, and a STAR agent within an MDP. In the figure, the human constructs a state s from the environment's display. In addition we assume that the human has both an effectiveness function (i.e., $H_e : S \times A \to \mathbb{R}$) and a social function (i.e., $H_s : S \times A \to \{0, 1\}$) as internal functions so that given a state s, and an action a that the agent has taken, the human is able to provide feedback to the agent that is consistent with them. The agent learns models of these two functions. Using the models, the agent's "effective-action selector" chooses an action which is then sent to the "social filter". If it passes the filter, in addition to it being performed, the action is also sent to the supervised learner along with the current state as an input. The supervised learner then refines the agent's models based on the information that this action is the most effective action among the permissible actions. Otherwise, the agent chooses a new (predicted to be less effective) action until it finds one that passes the social filter. Finally, we note that in STAR as in TAMER the learning is treated as a supervised learning problem, and does not require value propagation. This is due to the premise that humans provide feedback on the long-term effects of an action - the return, rather than the reward.

5 ACKNOWLEDGMENTS

This work has taken place in the Learning Agents Research Group (LARG) at UT Austin. LARG research is supported in part by NSF (IIS-1637736, IIS-1651089, IIS-1724157), ONR (N00014-18-2243), FLI (RFP2-000), DARPA, Intel, Raytheon, and Lockheed Martin. Peter Stone serves on the Board of Directors of Cogitai, Inc. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

REFERENCES

[1] S.V. Albrecht, J.W. Crandall, and S. Ramamoorthy. 2016. Belief and Truth in Hypothesised Behaviours. *Artificial Intelligence* 235 (2016), 63–94.

[2] Stefano V Albrecht and Subramanian Ramamoorthy. 2013. A game-theoretic model and best-response learning method for ad hoc coordination in multiagent systems. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1155–1156.

[3] Stefano V Albrecht and Peter Stone. 2017. Reasoning about hypothetical agent behaviours and their parameters. In *AAMAS '17*. 547–555.

[4] Thomas Arnold, Daniel Kasenberg, and Matthias Scheutz. 2017. Value alignment or misalignment–what will keep systems accountable. In *3rd International Workshop on AI, Ethics, and Society*.

[5] Avinash Balakrishnan, Djallel Bouneffouf, Nicholas Mattei, and Francesca Rossi. 2018. Using Contextual Bandits with Behavioral Constraints for Constrained Online Movie Recommendation.. In *IJCAI*. 5802–5804.

[6] Avinash Balakrishnan, Djallel Bouneffouf, Nicholas Mattei, and Francesca Rossi. 2019. Incorporating Behavioral Constraints in Online AI Systems. *n Proc. of the 33rd AAAI Conference on Artificial Intelligence (AAAI)* (2019).

[7] Samuel Barrett and Peter Stone. 2012. An Analysis Framework for Ad Hoc Teamwork Tasks. In *AAMAS '12*.

[8] S. Barrett and P. Stone. 2015. Cooperating with unknown teammates in complex domains: a robot soccer case study of ad hoc teamwork. In *AAAI-15*. 2010–2016.

[9] M. Chandrasekaran, P. Doshi, Y. Zeng, and Y. Chen. 2014. Team behavior in interactive dynamic influence diagrams with applications to ad hoc teams. In *AAMAS'14*. 1559–1560.

[10] Nicolas Cointe, Grégory Bonnet, and Olivier Boissier. 2016. Ethical judgment of agents' behaviors in multi-agent systems. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1106–1114.

[11] Joshua Greene, Francesca Rossi, John Tasioulas, Kristen Brent Venable, and Brian Charles Williams. 2016. Embedding Ethical Principles in Collective Decision Support Systems.. In *AAAI*, Vol. 16. 4147–4151.

[12] Dylan Hadfield-Menell and Gillian K Hadfield. 2018. Incomplete Contracting and AI Alignment. (2018).

[13] Dylan Hadfield-Menell, Smitha Milli, Pieter Abbeel, Stuart J Russell, and Anca Dragan. 2017. Inverse reward design. In *Advances in Neural Information Processing Systems*. 6768–6777.

[14] Dylan Hadfield-Menell, Stuart J Russell, Pieter Abbeel, and Anca Dragan. 2016. Cooperative inverse reinforcement learning. In *Advances in neural information processing systems*. 3909–3917.

[15] W Bradley Knox and Peter Stone. 2009. Interactively shaping agents via human reinforcement: The TAMER framework. In *the fifth international conference on Knowledge capture*. ACM, 9–16.

[16] Patrick MacAlpine and Peter Stone. 2017. Evaluating Ad Hoc Teamwork Performance in Drop-In Player Challenges. In *AAMAS'17*. Springer, 168–186.

[17] Francesca Rossi and Nicholas Mattei. 2018. Building Ethically Bounded AI. *arXiv preprint arXiv:1812.03980* (2018).

[18] Stuart Russell and Peter Norvig. 2010. *Intelligence artificielle: Avec plus de 500 exercices*. Pearson Education France.

[19] Amartya Sen and S Körner. 1974. Choice, ordering and morality. *Practical reason* (1974).

[20] Wendell Wallach and Colin Allen. 2008. *Moral machines: Teaching robots right from wrong*. Oxford University Press.

[21] Feng Wu, Shlomo Zilberstein, and Xiaoping Chen. 2011. Online Planning for Ad Hoc Autonomous Agent Teams. In *IJCAI*.

[22] Han Yu, Zhiqi Shen, Chunyan Miao, Cyril Leung, Victor R Lesser, and Qiang Yang. 2018. Building ethics into artificial intelligence. *arXiv preprint arXiv:1812.02953* (2018).