Implicitly Cooperative Agents through Impact-Aware Learning

David Rother[a;*]

[a]TU Darmstadt

Abstract. This research explores how autonomous agents learn and interact in shared environments, emphasizing the understanding of others as explicit agents rather than simple dynamic obstacles. When deploying robots in human-inhabited environments in the future, it will be unlikely that all interactions fit a predefined model of collaboration, where collaborative behavior is still expected from the robot. Utilizing the "theory of mind" concept, the research aims to infer the beliefs, policies, intentions, and goals of other agents, enabling the evaluation of our agent's impact on them. The study aims to create a multi-agent system capable of promoting inherent cooperation even with mixed objectives and adapting to various applications. Using Reinforcement Learning we develop a modular system that is capable to adapt to changing team sizes and motives for different agents. The developed method is trialed in a real-world assistant robot setup, testing cooperative actions without explicit initiation. Further evaluations occur in simulated environments, i.e. a cooking environment, to manage the policies of other agents and action recognition issues. We can measure the success of our method through the increased utility of either the population or single agents. Additionally, user studies can be conducted in which we can directly measure the satisfaction of humans when working alongside our agents and compare those to other methods.

1 Exploring the Potential for Inherent Cooperation in Multi-Agent Systems

Robotics research has focused on advancing the functional capabilities of robots but interacting with humans and other agents remains a hard challenge, especially in non-cooperative environments. In a shared environment, one or more additional agents may be present and change the environment drastically. All entities have some internal hidden state, that underlies their actions. To allow a robot to generate meaningful behavior in interactive environments, especially for interactions with humans, it is necessary to formally integrate models for these hidden states into the robots' planning methods. Earlier research focused on cooperative settings when an agent dealt with unknown teammates. This scenario is also widely known as ad-hoc teamwork [2]. The scope of this PhD is to research the learning and interaction of autonomous agents in environments, which are shared with others. In contrast to most state-of-the-art approaches, we want to treat others as explicit agents instead of simplified "dynamic obstacles" to enable cooperative and social behaviors. We want to approach this by inferring the beliefs, policies, intentions, or goals of the other agents and use this to evaluate the interaction impact of our agent's behaviors. To be able to do this, we need a model of the other agent, which we can feed with one of the inferred hidden parameters and then simulate the impact of different actions of the ego agent. Our approach is based on using the agent's own behavior/observation/world models and mapping them to the perspective of the interaction partner following the "theory of mind" concept [3]. Using such models, we can then evaluate the impact of our potential actions on the other agent's outcomes by running them on hypothetical future environments that incorporate the expected results of each of those actions. A related approach was used in a previous project [6] to learn the physical human-robot collaboration task of box turning. The robot predicted the future poses of the human given his own actions and evaluated the ergonomic quality of it to preferably select those with a high score. If the agent's reward depends, directly or indirectly, on the future rewards of others, this would allow choosing an interaction strategy that minimizes the potential losses of a partner while still leading to the achievement of its own target. We want to extend the agent's capability to use a theory of mind of others to play through hypothetical scenarios in his head and use the information about others to learn faster about previously unseen states. The resulting behavior should lead to a socially aware multi-agent system with inherent cooperation even in the case of non-overlapping targets or mixed objectives. As this approach uses decentralized agents, unlike many of the related works, it also allows one to interact with novel agents that are not part of the learning, in particular with humans. We target to develop a method and system design that generalizes for different applications, given the assumption that the agent learned/knows how to solve certain tasks that might be performed by a human. We aim at working in situations, where a human performs various tasks, and the robot should be able to incorporate the outcome of his own actions with respect to the human task performance. In such a way, the robot detects opportunities for cooperative actions without having an explicit task assignment. So far similar applications were limited to planning execution orders in joint tasks [5]. For example, in work by [1] a robot was evaluating a joint task state and either initiated its own action whenever a task-relevant was available (i.e., not performed by the human already) or only when detecting a lack of progress at the human side and showed that the former leads to better objective and subjective results. However, in all those cases the robot's target is resolving the task rather than helping the human resolve a task, which has implications for the generalization of tasks.

2 Contribution outline

First, the agent needs a general learning scheme for solving a given task, where the agent weighs its own goals with the other agent's

* Corresponding Author. Email: david.rother@tu-darmstadt.de

needs. This is approached using methods of Reinforcement Learning and specifically maximum entropy methods. We employ state-of-the-art approaches, e.g. using agent self-play or combining model-based with model-free aspects. Second, the methods used by our agent must allow us to model other agents' policies or Value functions which should be used as a modulator for the reward function. Third, to be able to use the learned model for other agents, there needs to be a means to infer the goals (and possibly other things like capabilities) of the interaction partners. Recognizing goals or intentions is a major research field. We rely on fixed assumptions or, when possible, existing algorithms. Nevertheless, our method is able to incorporate uncertainties of the goal estimates. Forth, our system needs a knowledge representation of the domain he is in, including knowledge about tasks, the environment, and the capabilities of other agents. In our first work, we assume the same capabilities of all agents, knowledge of possible tasks, and a fully observable environment. A target application to test the developed method in the real world is an assistant robot in an apartment setup. In a possible realization, the robot knows for a number of tasks, which objects are used and in which order or places. When interacting with objects, the system is able to integrate the potential impact of certain object (dis-)placements on current or potential human tasks in its decision-making. This should lead to a robot that performs cooperative actions without the need for an explicit initiation or even an explicit joint task. An example could be joint meal preparation, where the socially aware robot understands the the current state of the human in the task (e.g. put pot with water on the stove) and performs an action that brings the human to an even better state (e.g. hand him salt). This is supported by the development of the overcooked environment and implementation of various learning algorithms. We also plan to compile a set of multi-agent environments where we can evaluate the framework in diverse settings and propose means to measure the performance of both robots and humans. We will incorporate settings where other agents follow a common goal (cooperation), or independent goals (coexist), or might have conflicting goals (game/competition). First analyses and evaluations are done in simulated environments (e.g. Overcooked Environment, explicit simulation of robots) in order to be able to control the policies of the other agents and to abstract from the issues of sensor noise and action recognition.

3 Existing Contributions

Towards the mentioned goals two major contributions have been made, that will be published at ECAI 2023 [4]. First, a new cooking simulation environment is published [1], inspired by the game Overcooked as well as previous work using another cooking environment [7]. The previous work was limited to a single joint goal, modeling only strictly cooperative tasks, whereas our environment allows each agent to have separate recipes, modeling the complex space of mixed-motive environments. We offer a fixed vector length input representation of the state space between different layouts of the environment. Our environment supports the latest versions of the gymnasium [2] and pettingzoo [3] libraries, enabling easy usage of common RL frameworks. We support a reward scheme with rewards for subgoals and only for complete dishes as well as a configurable time penalty. By offering this simulation environment future research can be compared in more complex scenarios.

[1] https://github.com/DavidRother/cooking_zoo
[2] https://gymnasium.farama.org/
[3] https://pettingzoo.farama.org/

Second, we developed a new way to make decisions, when there is an unknown number of agents in a mixed-motive scenario. Following, the novelty of the framework is briefly described. Our framework trains task and interaction policies. Task policies can be trained alone, whereas interaction policies are trained with another agent by estimating the impact one has on its expected reward. We compute the Q-values of the given task and the interactions for each action. We obtain the weighting of Q-values using the entropy of the task distribution and the Jensen-Shannon-Distance between the task action distribution and the impact action distributions. We were able to show that the entropy-based blending mechanism is within a certain error bound of the optimal policy combination w.r.t. the equal-weighted sum of rewards of the agents and that this bound is minimal for maximum entropy policies.

Our methods show an improvement in the overall utility of groups of agents as well as retaining high individual scores for socially capable agents, indicating that they are able to strike a compromise in trading off their own reward and others while retaining their goal-reaching capabilities.

4 Impact and Future Work

In future work, we plan to test our setup directly with human users and evaluate their sentiment toward new socially capable agents through questionnaires in a simulator and on a real demonstration with a robot. Furthermore, we develop an algorithm to approximate and learn the best combination of policies on the fly to be able to reflect the wishes and preferences of users. There the agent does not need to solve a task given beforehand but instead optimizes the group reward of users as the primary optimization objective.

We believe our research will be crucial to scaling multi-agent applications to real decision-making with arbitrary tasks with an unknown number of agents. Currently, the work has limitations in that learning interaction policies is still hard as the credit of one's own action towards others' success is not trivial to compute and we are currently limited to cooperation in tasks that our agent has previously seen and trained on.

References

[1] Jimmy Baraglia, Maya Cakmak, Yukie Nagai, Rajesh PN Rao, and Minoru Asada, 'Efficient human-robot collaboration: when should a robot take initiative?', *The International Journal of Robotics Research*, **36**(5-7), 563–579, (2017).

[2] Reuth Mirsky, Ignacio Carlucho, Arrasy Rahman, Elliot Fosong, William Macke, Mohan Sridharan, Peter Stone, and Stefano V Albrecht, 'A survey of ad hoc teamwork research', in *European Conference on Multi-Agent Systems (EUMAS)*, (2022).

[3] Neil Rabinowitz, Frank Perbet, Francis Song, Chiyuan Zhang, SM Ali Eslami, and Matthew Botvinick, 'Machine theory of mind', in *International conference on machine learning*, pp. 4218–4227. PMLR, (2018).

[4] David Rother, Thomas Weisswange, and Jan Peters, 'Disentangling interaction using maximumentropy reinforcement learning in multi-agent systems', in *European Conference on Artificial Intelligence*, (2023).

[5] Panagiota Tsarouchi, Sotiris Makris, and George Chryssolouris, 'Human–robot interaction review and challenges on task planning and programming', *International Journal of Computer Integrated Manufacturing*, **29**(8), 916–931, (2016).

[6] Linda van der Spaa, Michael Gienger, Tamas Bates, and Jens Kober, 'Predicting and optimizing ergonomics in physical human-robot cooperation tasks', in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1799–1805. IEEE, (2020).

[7] Sarah A. Wu, Rose E. Wang, James A. Evans, Joshua B. Tenenbaum, David C. Parkes, and Max Kleiman-Weiner, 'Too many cooks: Coordinating multi-agent collaboration through inverse planning', *Topics in Cognitive Science*, **n/a**(n/a), (2021).