Hierarchical Reinforcement Learning for Ad Hoc Teaming

Extended Abstract

Stéphane Aroca-Ouellette
University of Colorado Boulder
Boulder, CO, USA

Miguel Aroca-Ouellette
London, United Kingdom

Upasana Biswas
Arizona State University
Phoenix, AZ, USA

Katharina Kann
University of Colorado Boulder
Boulder, CO, USA

Alessandro Roncone
University of Colorado Boulder
Boulder, CO, USA

ABSTRACT

When performing collaborative tasks with new unknown teammates, humans are particularly adept at adapting to their collaborator and converging toward an aligned strategy. However, state of the art autonomous agents still do not have this capability. We propose that a critical reason for this disconnect is that there is an inherent hierarchical structure to human behavior that current agents lack. In this paper, we explore the use of hierarchical reinforcement learning to train an agent that can navigate the complexities of ad hoc teaming at the same level of abstraction as humans. Our results demonstrate that when paired with humans, our Hierarchical Ad Hoc Agent (HAHA) outperforms all baselines on both the team's objective performance and the human's perception of the agent.

KEYWORDS

Human Agent Collaboration; Ad Hoc Teaming; Zero-Shot Coordination; Mutual Adaptation; Hierarchical Reinforcement Learning;

ACM Reference Format:
Stéphane Aroca-Ouellette, Miguel Aroca-Ouellette, Upasana Biswas, Katharina Kann, and Alessandro Roncone. 2023. Hierarchical Reinforcement Learning for Ad Hoc Teaming : Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

Increasingly, the AI community has recognized the importance of creating agents that can collaborate effectively with humans. Drastic variance in human knowledge, ability, and preference require adaptive agents that can quickly conform to new and unseen teammates—an ability frequently named ad hoc teaming [2] or zero-shot coordination [4]. Previous work has made valuable strides toward this goal by investing the teammate(s) that agents are trained with. [3] proposes behavioral cloning play (BCP), where an agent trains with a behavioral cloning teammate; [5, 10] propose training with a set of teammates that vary in skill level. However, this previous work has focused on agents that learn low-level policies (e.g. move *LEFT, RIGHT, UP, DOWN,* ..). Notably, there is an inherent hierarchical structure to human behavior [12] that these techniques fail to capture. Humans utilize task hierarchies to analyze and manage projects [1, 13] and synchronize with other humans [6, 8, 9].

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

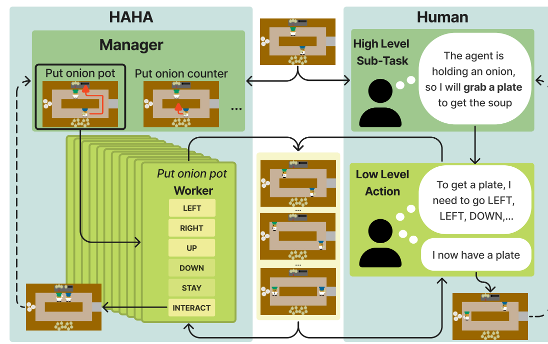

Figure 1: An overview of HAHA. Given the current state, the manager decides to place its onion into the pot and invokes the correct worker to carry out the low-level actions. Concurrently, the human notices that the agent is focusing on onions, so decides to grab a plate to serve the completed soup.

In this work, **we hypothesize that providing an autonomous agent with a hierarchical structure will cognitively and behaviorally align the agent to humans, facilitating human-agent teams to understand each other and adapt to aligned strategies**. To this end, we explore the use hierarchical reinforcement learning (HRL) to train an agent that navigates the complexities of ad hoc teaming at the same level of abstraction as humans.

To our knowledge, we are the first paper to propose, motivate, and demonstrate the benefit of using HRL for human agent interaction. We propose the following novel contributions: 1) we present our Hierarchical Ad Hoc Agent (HAHA) capable of zero-shot coordination with humans; 2) we demonstrate that HAHA outperforms all baselines when paired with unseen autonomous agents; 3) we demonstrate that HAHA outperforms and is preferred over baselines when paired with humans.

2 HAHA: HIERARCHICAL AD HOC AGENT

Following prior work in ad hoc teaming [3, 10], we study the use of our hierarchical approach using a simplified Overcooked environment, as proposed in [3]. In this work, we demonstrate significant improvement using only three of the five original layouts (cf. Fig. 2).

We start with the motivation of aligning agents and humans at a cognitive and behavioral level by using a hierarchical structure—with high level sub-tasks (e.g. put an onion in a pot, or serving a soup, ...) that organize low-level actions (e.g. *UP, DOWN, LEFT, RIGHT, INTERACT*). To achieve this, we leverage FuNs [11] to develop HAHA: Hierarchical Ad Hoc Agents.

Layout	SP	BCP	FCP	HAHA
AA	193.7 (83.1)	194.0 (83.3)	**233.0 (100)**	215.3 (92.4)
CC	45.0 (62.8)	6.0 (8.4)	51.0 (71.1)	**71.7 (100)**
FC	52.3 (84.4)	**62.0 (100)**	31.7 (51.1)	60.3 (97.3)
Avg.	97.0 (76.7)	87.3 (63.9)	105.2 (74.1)	**115.8 (96.5)**

Table 1: Average performance across ten trials of HAHA and the FCP, BCP, and SP baselines when paired with unseen agents. Normalized scores are in parentheses. Each trial is run for 80 seconds (T=400 steps).

The environment's *INTERACT* action is used to pick up and put down objects, which we are able to use to delineate a set of twelve high-level sub-tasks that covers all movement patterns. Twelve worker agents are then trained using PPO [7] to complete each of the sub-tasks. The worker's action space is the environment's low-level actions, and they receive a positive of negative reward depending on whether they successfully complete the sub-task or not. Lastly, a manager is trained using PPO to identify which sub-task should be performed next. Specifically, a manager's action space is picking which worker will control the agent until the next *INTERACT* action is performed, and receives the environment's normal rewards. Fig. 1 shows an overview of our method. Code is available at: https://github.com/HIRO-group/HAHA.

3 RESULTS

Ad hoc teaming requires mutual adaptation: being able to adapt oneself and making it easy for a teammate to adapt to you. While playing with unseen agents tests the former, only experiments with real humans capture the full challenge of ad hoc teaming. To this end, we evaluate HAHA against three baselines—self-play (SP), behavioral cloning play (BCP) [3], and Fictitious Co-Play (FCP) [10]—when paired both with previously unseen agents and with humans through an online user study ($n = 55$).

3.1 Ad Hoc Teaming with Unseen Agents

We first have the agents play with unseen agents: they play with each other, with themselves, with a human proxy, and with a random agent. The results are shown in Table 1. In this set up, the only agent who has trained with any of the teammates before is the SP agent, who was trained with itself. The results show that HAHA performs 20.00% better than the average of the baselines, and 10.07% than the second best agent.

3.2 Ad Hoc Teaming with Humans

Result in Fig. 2 show that HAHA significantly outperforms all baselines when paired with humans. It outperforms the average of the baselines by 31.89% and is 27.79% better than the second best agent. This further improvement over the baselines when paired with humans strongly supports our motivation that human-agent collaboration benefits from aligning humans and agents at a cognitive and behavioral level. We hypothesize that part of this improvement is that HAHA is always task oriented, making it easier for the human teammates to understand what HAHA is trying to accomplish. The results of our likert scale questions, shown in Fig. 3, supports this claim, with humans finding HAHA significantly more understandable than the baselines. Humans also found HAHA to be significantly more fluent, trustworthy, and helped them better adapt to the task.

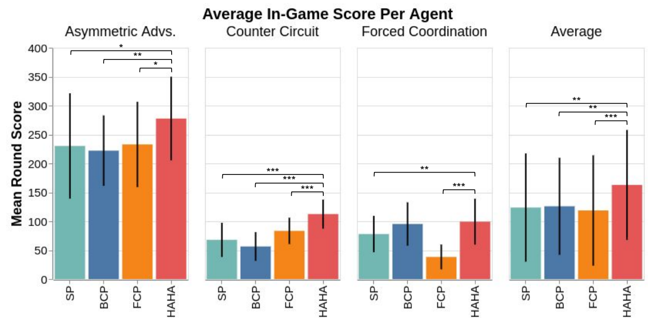

Figure 2: Average reward of HAHA and the SP, BCP, and FCP when paired with humans on each layout. Asterisks indicate significance: *=(p<0.05), **=(p<0.005), *=(p<0.0005).**

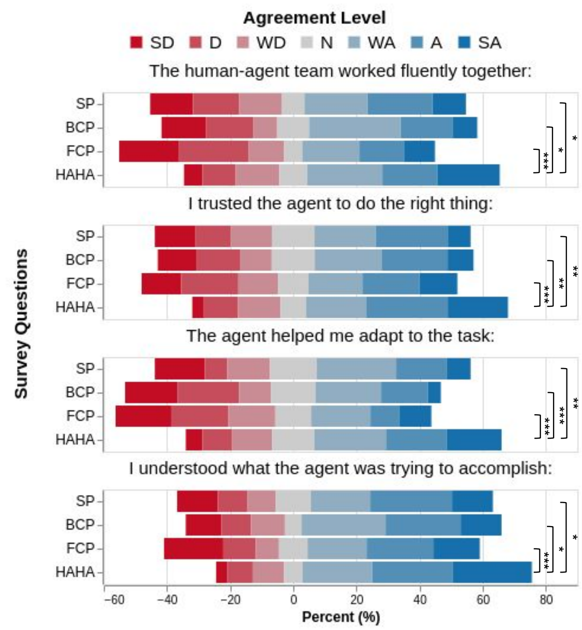

Figure 3: Results from the likert questions. Bluer/Redder bars indicate that higher agreement/disagreement. Asterisks indicate significance: *=(p<0.05), **=(p<0.005), *=(p<0.0005). SD=Strongly Disagree, D=Disagree, WD=Weakly Disagree, N=Neutral, WA=Weakly Agree, A=Agree, SA=Strongly Agree**

4 CONCLUSION

In this paper, we have investigated the use of HRL for effective human-agent collaboration. The experiments we conducted demonstrates that when paired with either other unseen agents, or with real humans, our Hierarchical Ad Hoc Agent significantly and substantially outperforms all baselines.

Lastly we note that because the proposed architecture is independent to the training teammates, the work presented is orthogonal and complementary to all previous work and future work in the Overcooked environment that investigates training teammates.

ACKNOWLEDGMENTS

This work was supported by the Army Research Laboratory under Grants W911NF-21-2-02905 and W911NF-21-2-0126 and by the Office of Naval Research under Grant N00014-22-1-2482.

REFERENCES

[1] John Annett. 2003. Hierarchical task analysis. *Handbook of cognitive task design* 2 (2003), 17–35.

[2] Samuel Barrett, Avi Rosenfeld, Sarit Kraus, and Peter Stone. 2016. Making Friends on the Fly: Cooperating with New Teammates. *Artificial Intelligence* (October 2016). https://doi.org/10.1016/j.artint.2016.10.005

[3] Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. 2019. On the Utility of Learning about Humans for Human-AI Coordination. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2019/file/f5b1b89d98b7286673128a5fb112cb9a-Paper.pdf

[4] Hengyuan Hu, Adam Lerer, Alex Peysakhovich, and Jakob Foerster. 2020. "Other-Play" for Zero-Shot Coordination. In *Proceedings of the 37th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 119)*, Hal Daumé III and Aarti Singh (Eds.). PMLR, 4399–4410. https://proceedings.mlr.press/v119/hu20a.html

[5] Keane Lucas and Ross E. Allen. 2022. Any-Play: An Intrinsic Augmentation for Zero-Shot Coordination. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems* (Virtual Event, New Zealand) *(AAMAS '22)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 853–861.

[6] Olivier Mangin, Alessandro Roncone, and Brian Scassellati. 2022. How to be Helpful? Supportive Behaviors and Personalization for Human-Robot Collaboration. *Frontiers in Robotics and AI* 8 (2022), 426.

[7] Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. 2021. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research* 22, 268 (2021), 1–8. http://jmlr.org/papers/v22/20-1364.html

[8] Alessandro Roncone, Olivier Mangin, and Brian Scassellati. 2017. Transparent role assignment and task allocation in human robot collaboration. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 1014–1021. https://doi.org/10.1109/ICRA.2017.7989122

[9] Neville A Stanton. 2006. Hierarchical task analysis: Developments, applications, and extensions. *Applied ergonomics* 37, 1 (2006), 55–79.

[10] DJ Strouse, Kevin McKee, Matt Botvinick, Edward Hughes, and Richard Everett. 2021. Collaborating with Humans without Human Data. In *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (Eds.), Vol. 34. Curran Associates, Inc., 14502–14515. https://proceedings.neurips.cc/paper/2021/file/797134c3e42371bb4979a462eb2f042a-Paper.pdf

[11] Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David Silver, and Koray Kavukcuoglu. 2017. FeUdal Networks for Hierarchical Reinforcement Learning. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (Sydney, NSW, Australia) *(ICML'17)*. JMLR.org, 3540–3549.

[12] Gwydion Williams. 2022. *Hierarchical Influences on Human Decision-Making*. Ph.D. Dissertation. UCL (University College London).

[13] Yue Maggie Zhou. 2013. Designing for complexity: Using divisions and hierarchy to manage complex tasks. *Organization Science* 24, 2 (2013), 339–355.