Hiking up that HILL with Cogment-Verse: Train & Operate Multi-agent Systems Learning from Humans

Demonstration Track

Sai Krishna Gottipati
AI Redefined
Montréal, Canada
sai@ai-r.com

Luong-Ha Nguyen
AI Redefined
Montréal, Canada
ha@ai-r.com

Clodéric Mars
AI Redefined
Montréal, Canada
cloderic@ai-r.com

Matthew E. Taylor
AI Redefined
Montréal, Canada
matt@ai-r.com

ABSTRACT

As more AI systems are deployed, humans are increasingly required to interact with them in multiple settings. However, such AI systems seldom learn from these interactions with humans, which provides an important opportunity to improve from human expertise and context awareness. Several recent results in the fields of reinforcement learning (RL) and human-in-the-loop learning (HILL) show that AI agents *can* perform better when humans are involved in their training process. Humans can provide rewards to the agent, demonstrate tasks, design curricula, or act directly in the environment, but these potential performance improvements also come with architectural, functional design, and engineering complexities. This paper discusses Cogment, a unifying open-source framework that introduces a formalism to support a variety of human(s)-agent(s) collaboration topologies and training approaches. Cogment addresses the complexity of training with humans within a production-ready platform. On top of Cogment, we introduce Cogment Verse a research platform dedicated to the research community to facilitate the implementation of HILL and Multi-Agent RL experiments. With these platforms, our end goal is to enable the generalization of intelligence ecosystems where AI agents and humans learn from each other and collaborate to address increasingly complex or sensitive use cases. The video demonstration is available at https://youtu.be/v-K0DqIL9K0

KEYWORDS

Multi Agent RL; Human-in-the-loop; Open Source; Real world applications

ACM Reference Format:
Sai Krishna Gottipati, Luong-Ha Nguyen, Clodéric Mars, and Matthew E. Taylor. 2023. Hiking up that HILL with Cogment-Verse: Train & Operate Multi-agent Systems Learning from Humans: Demonstration Track. 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

The involvement of AI systems in many aspects of our societies, industries, and everyday life is increasing. It is also becoming increasingly clear that in order to ensure that their involvement is beneficial, the AI agents should be able to interact directly with the people they are designed to support, from training to operationalization and use.

Many systems are too critical to be fully trusted to be completely autonomous agents (e.g., medical applications), necessitating collaboration between humans and AI agents. In many other use cases, the complexity or lack of data renders traditional AI methods unable to provide full automation in a reasonable time frame, and collaborating and learning from human expertise then become a central aspect of such systems. Additionally, people operate at different speeds than AI systems, and their time is precious — it is critical to efficiently leverage what humans can contribute in such a collaboration paradigm. Addressing these constraints starts with an environment in which humans and AI agents can operate and train together.

In the past few years, human involvement has grown beyond data annotation to become what we call human-in-the-loop learning (HILL); providing AI agents with feedback and guidance. In many domains, the performance of AI agents was shown to improve by taking feedback, as reward or preferences, from humans [15, 25], by learning from human demonstrations [14], or taking advantage of human input in other ways, as discussed later. However, there was no unifying framework that allowed researchers to quickly develop applications that supported HILL or that enabled engineers to deploy at scale. Cogment is designed to address these needs. It is a framework that facilitates the development and deployment of projects involving multiple actors (humans or AI agents) that interact with each other in a simulated or real environment.

Cogment Verse is built upon Cogment [16], an open source platform designed to address the challenges of building and operating human-in-the-loop learning systems.

The first of those challenges is interoperability between simulations, deep learning or other AI frameworks, and interactive user interfaces. Thanks to its distributed micro service architecture, Cogment is able to execute episodes over any kind of simulation or

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.

agent decision making and involve any kind of interactive application.

The second challenge is the limited availability of humans. Unlike AI agents, humans are not available 100% of the time and cannot operate faster than real time. Thanks to its fully asynchronous orchestration, Cogment does not stop training while waiting for humans to be available or to decide on the next action. Because it can be deployed on the cloud at scale, Cogment is also well suited to leverage large crowds of humans.

The third challenge is the cost of acquiring human data. Because Cogment episodes are configuration-driven, it is easy to have a mixed training curriculum involving both interactive episodes and headless ones. Due to its unified data store, Cogment enables the training process to leverage online data as it is collected and also mix in valuable historical data.

2 COGMENT VERSE

To enable researchers to get started with Cogment easily, we introduce Cogment Verse (https://github.com/cogment/cogment-verse) which includes several code examples. They include off-policy RL algorithms like DQN [11], rainbow DQN [6], DDPG [8], TD3 [5], and on-policy algorithms like advantage actor-critic (A2C) [10] and proximal policy optimization (PPO) [17]. Cogment Verse includes base implementations of these algorithms in both single and multi-agent settings. Different kinds of HILL paradigms like learning from demonstrations (using behavioral cloning [19]), learning from human interventions [3], and explicit human feedback [21] are also included. In all these examples, multiple humans can interact with one or multiple learning agents simultaneously. These RL, MARL, and HILL algorithms are tested with a wide variety of environments including simple OpenAI Gym environments [2], Atari, MinAtar [23], boardgames, card games, and other multi-agent environments in PettingZoo [18], procedural generation environments (i.e., Procgen [4]), and robotics environments such as Robosuite [26] and IsaacGym [9].

Different research groups have used Cogment & Cogment Verse in a wide range of applications, which we discuss in the following sections.

3 HUMAN-MACHINE TEAMING FOR AIR DEFENSE

Defense-critical applications, such as securing airspace from intrusions, are of paramount importance. Complete automation of such a defense system is impossible owing to the potential real-world impact. On the other hand, continuous control and monitoring of such systems by humans is infeasible as well because of the amount of continual oversight needed. We developed a system where humans and embodied AI agents can collaborate towards a successful defense of the airport's air space.

A team of five ally drones was tasked with protecting an airspace against one or more enemy drones. The complete experimental setup was implemented in Cogment. The ally drones were pre-trained from human demonstrations (collected from multiple non-expert humans). They were then trained in a standard RL setting using D3QN [20]. The policy is also updated on the basis of human intervention and feedback at any time during the episode.

4 MULTI-TEACHER ASYMMETRIC SELF-PLAY

Training a goal-conditioned agent that can generalize to unseen goals in sparse reward environments is a difficult challenge. Plappert et al. [13] proposed to tackle this by introducing a 'teacher' agent that proposes increasingly difficult goals (e.g., via a curriculum) to the goal-conditioned 'student' agent. We further extended this idea [7] by using multiple teachers to increase the diversity of the goals generated and improve the training speed and sample complexity. Using Cogment allowed us to switch between actors (student and the teacher agents), use each other's replay experiences, and run multiple parallel trials were leveraged in this setting. The code is also open-sourced at https://github.com/kharyal/selfplayRL/tree/fetch_reach. Another ongoing research project is utilizing Cogment for learning to reach goals from natural language instructions — also within the multi-teacher asymmetric self-play framework.

5 TIMBER HARVESTING

In [24], the authors proposed a system to operate a complex machine for timber harvesting. For further experimentation involving learning from human demonstrations, the authors are utilizing Cogment for simultaneous multiple human demonstrations and potentially for learning from human feedback as the reward signals in such environments are sparse.

6 WARM START OFF-POLICY RL

Cogment has been used [22] for analyzing the behavior of RL policies that are pre-trained using methods like behavior cloning. The authors also compared with the off-policy setting and proposed a novel method, Confidence Constrained Learning (CCL), for warm starting RL. CCL improves learning by balancing between the policy gradient and constrained learning according to a confidence measure of the Q-values. The offline and off-policy learning capabilities of Cogment via its data store were leveraged in this work.

7 HANABI

Hanabi was originally proposed as a benchmark for training cooperative agents [1]. Later, it was used [12] to measure the zero-shot training capabilities in collaborating with novel players. As part of ongoing research, Cogment Verse implemented the entire pipeline of 1) self-play, 2) training together with randomly selected pre-trained agents, 3) testing and improving zero-shot coordination capabilities by involving human players during training or testing phases [12]. The ability of Cogment to efficiently switch between different actors, run multiple parallel trials, and support different training paradigms (self-play in phase-1 and multi-agent in phase-2) simultaneously helped in this endeavor.

8 CONCLUSION

This paper has given a high-level introduction to Cogment and Cogment Verse. Our hope is that readers will consider using this open-source framework for their own research in HILL, multi-agent simulations, or reinforcement learning. This framework will allow researchers to quickly test ideas, scale to large numbers of humans and agents in a single system, and integrate high-fidelity simulations. We also welcome ideas for improvement or collaboration.

REFERENCES

[1] Nolan Bard, Jakob N. Foerster, Sarath Chandar, Neil Burch, Marc Lanctot, H. Francis Song, Emilio Parisotto, Vincent Dumoulin, Subhodeep Moitra, Edward Hughes, Iain Dunning, Shibl Mourad, Hugo Larochelle, Marc G. Bellemare, and Michael Bowling. 2020. The Hanabi challenge: A new frontier for AI research. *Artif. Intell.* 280 (March 2020), 103216. https://doi.org/10.1016/j.artint.2019.103216

[2] Greg Brockman. 2016. OpenAI Gym. (2016). arXiv:1606.01540 http://arxiv.org/abs/1606.01540

[3] Sonia Chernova and Manuela Veloso. 2009. Interactive Policy Learning through Confidence-Based Autonomy. *J. Artif. Int. Res.* 34, 1 (jan 2009), 1–25.

[4] Karl Cobbe, Christopher Hesse, Jacob Hilton, and John Schulman. 2019. Leveraging Procedural Generation to Benchmark Reinforcement Learning. *arXiv preprint arXiv:1912.01588* (2019).

[5] Scott Fujimoto, Herke van Hoof, and David Meger. 2018. Addressing Function Approximation Error in Actor-Critic Methods. *ArXiv* abs/1802.09477 (2018).

[6] Matteo Hessel. 2018. Rainbow: Combining improvements in deep reinforcement learning. In *Thirty-second AAAI conference on artificial intelligence*.

[7] Chaitanya Kharyal, Tanmay Kumar Sinha, SaiKrishna Gottipati, Srijita Das, and Matthew E. Taylor. 2022. Do As You Teach: A Multi-Teacher Approach to Self-Play in Deep Reinforcement Learning. In *Deep Reinforcement Learning Workshop NeurIPS 2022*. https://openreview.net/forum?id=KEH4KSoJh2W

[8] Timothy Lillicrap, Jonathan Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2015. Continuous control with deep reinforcement learning. *CoRR* (09 2015).

[9] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, and Gavriel State. 2021. Isaac Gym: High Performance GPU-Based Physics Simulation For Robot Learning.

[10] Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Tim Harley, Timothy P. Lillicrap, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous Methods for Deep Reinforcement Learning. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48* (New York, NY, USA) *(ICML'16)*. JMLR.org, 1928–1937.

[11] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. 2013. Playing Atari with Deep Reinforcement Learning. *ArXiv* abs/1312.5602 (2013).

[12] Hadi Nekoei, Akilesh Badrinaaraayanan, Aaron C. Courville, and Sarath Chandar. 2021. Continuous Coordination As a Realistic Scenario for Lifelong Learning. *CoRR* abs/2103.03216 (2021). arXiv:2103.03216 https://arxiv.org/abs/2103.03216

[13] OpenAI OpenAI, Matthias Plappert, Raul Sampedro, Tao Xu, Ilge Akkaya, Vineet Kosaraju, Peter Welinder, Ruben D'Sa, Arthur Petron, Henrique Ponde de Oliveira Pinto, Alex Paino, Hyeonwoo Noh, Lilian Weng, Qiming Yuan, Casey Chu, and Wojciech Zaremba. 2021. Asymmetric self-play for automatic goal discovery in robotic manipulation. https://openreview.net/forum?id=hu2aMLzOxC

[14] Takayuki Osa. 2018. An Algorithmic Perspective on Imitation Learning. abs/1811.06711 (2018). arXiv:1811.06711

[15] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155* (2022).

[16] A. I. Redefined, Sai Krishna Gottipati, Sagar Kurandwad, Clodéric Mars, Gregory Szriftgiser, and François Chabot. 2021. Cogment: Open Source Framework For Distributed Multi-actor Training, Deployment & Operations. *CoRR* abs/2106.11345 (2021). arXiv:2106.11345 https://arxiv.org/abs/2106.11345

[17] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. *CoRR* abs/1707.06347 (2017). arXiv:1707.06347 http://arxiv.org/abs/1707.06347

[18] Justin K. Terry, Benjamin Black, Ananth Hari, Luis Paulo Santos, Clemens Dieffendahl, Niall L. Williams, Yashas Lokesh, Caroline Horsch, and Praveen Ravi. 2020. PettingZoo: Gym for Multi-Agent Reinforcement Learning. In *Neural Information Processing Systems*.

[19] Faraz Torabi, Garrett Warnell, and Peter Stone. 2018. Behavioral Cloning from Observation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence* (Stockholm, Sweden) *(IJCAI'18)*. AAAI Press, 4950–4957.

[20] Hado Van Hasselt, Arthur Guez, and David Silver. 2016. Deep reinforcement learning with double q-learning. In *AAAI*.

[21] Garrett Warnell, Nicholas R. Waytowich, Vernon Lawhern, and Peter Stone. 2017. Deep TAMER: Interactive Agent Shaping in High-Dimensional State Spaces. *CoRR* abs/1709.10163 (2017). arXiv:1709.10163 http://arxiv.org/abs/1709.10163

[22] Benjamin Wexler, Elad Sarafian, and Sarit Kraus. 2022. Analyzing and Overcoming Degradation in Warm-Start Reinforcement Learning. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 4048–4055. https://doi.org/10.1109/IROS47612.2022.9981286

[23] Kenny Young. 2019. MinAtar: An Atari-inspired Testbed for More Efficient Reinforcement Learning Experiments. (2019). arXiv:1903.03176 http://arxiv.org/abs/1903.03176

[24] Ehsan Yousefi, Dylan P. Losey, and Inna Sharf. 2022. Assisting Operators of Articulated Machinery with Optimal Planning and Goal Inference. In *2022 International Conference on Robotics and Automation (ICRA)*. 2832–2838. https://doi.org/10.1109/ICRA46639.2022.9811864

[25] Ruohan Zhang. 2019. Leveraging Human Guidance for Deep Reinforcement Learning Tasks. International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2019/884

[26] Yuke Zhu, Josiah Wong, Ajay Mandlekar, Roberto Martín-Martín, Abhishek Joshi, Soroush Nasiriany, and Yifeng Zhu. 2020. robosuite: A Modular Simulation Framework and Benchmark for Robot Learning. In *arXiv preprint arXiv:2009.12293*.