8-19-2021

Multi-Agent Radiation Localization

Teresa Nguyen
Portland State University

Multi-Agent Radiation Localization

Teresa Nguyen (*Author*)
Computational Modeling: Serving the City
REU at Portland State University
Portland, OR USA
teresa.nguyen2@pcc.edu

***Abstract*— Advancement of radiation detection technology is an ongoing process, and adjustments are made based on pre-existing conditions of radiation presence--both natural and man made. Tools that are currently used for safely detecting radiation in urban environments exist in several forms: drones, robots, or handheld radiation detection devices. This is a harm reductive way to explore radiation-infected environments while preserving human health as best as possible. In order for these autonomous platforms to successfully detect radiation sources, an algorithm needs to be created that is capable of gathering crucial data on its own with little to no human interference. Machine learning has been the algorithm of choice for researchers, particularly reinforcement learning and deep reinforcement learning. These tools for information gathering are designed to have the algorithm "learn" on its own based on a reward system that allows for seekers called "agents" to fulfill its mission objective of radiation detection in urban environments. In this paper, we explore the capability of having multiple agents within a controlled environment learn to locate a radiation source on their own using a tactic called Differentiable Inter-Agent Learning. This concept would be built upon pre-existing work developed in a master's thesis examining Proximal Policy Optimization for radiation source search using a single agent. By adding what could potentially be many agents to a radiation detection algorithm, it could provide a quicker and more efficient strategy for locating radiation sources remotely to preserve human health and safety. This multi-agent algorithm would examine if this is possible.**

Keywords—Research Experience for Undergraduates (REU), Multi-Agent Radiation Localization, Reinforcement Learning (RL), Multi-Agent Reinforcement Learning (MARL), Deep Reinforcement Learning (DRL), Proximal Policy Optimization (PPO), Multi-agent System (MAS), Differentiable Inter-agent Learning (DIAL)

I. Introduction

Human advancement of nuclear technology has created opportunities for scientists and civilian populations to benefit greatly from energy production and medical applications. With each new discovery made in the name of progress comes an equal amount (if not more) of challenges to maintain human safety [10]. One of these challenges involves the risk of radiation exposure, and the solution finding methods involved in successfully detecting radiation sources.

A. Background

The ability to detect, localize, and identify these sources are dependent upon things like measured gamma-ray spectrum from a radiation detector [7]. The severity of radiation exposure is probabilistic in nature, as there is also a decaying factor involved as well. Exposure ranges from natural existence of radiation (the sun, outer space) to man-made (radiation in medicine). [10]. A way to investigate the success rate of a radiation detector is to set up a simulated environment allowing

for an agent to work through training data and locate radiation sources in a more efficient manner. This strategy was done by Philippe Proctor within the last year, where a single agent was trained using Proximal Policy Optimization (PPO) for Radiation Source Search [7].

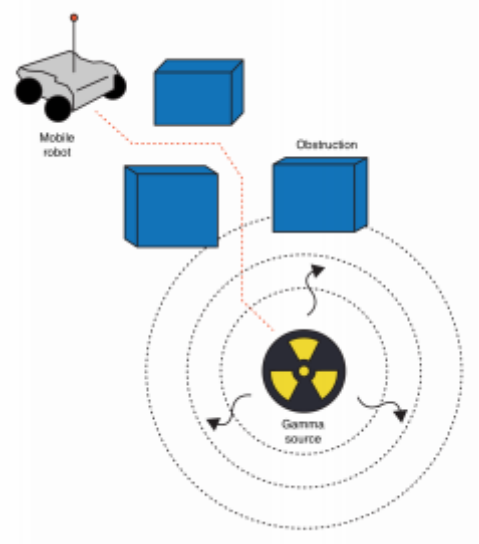

Fig 1. : Autonomous mobile robot moving through a non-convex (with obstacles) local environment to detect a gamma radiation source [7]

PPO is a strategy used in Reinforcement Learning (RL) to allow for optimal performance of the agents being assigned to the task at hand. [6] It is recognized as DRL using an on-policy, model-free, stochastic gradient that can be used to try and predict where the detector thinks a radiation source might be. [7] Efficiency can be further realized using Multi-Agent Based Radiation Localization, a tactic building off of Proctor's single agent PPO algorithm. This type of localization will use Proctor's discrete environment setup with controlled parameters to manage computational cost for training data. The goal of multi-agent PPO performance is to gather information more quickly from multiple radiation detectors so that automated platforms such as drones or robots can enter the field to detect radiation levels instead of humans. [6] This preserves human health from radiation exposure and allows for more efficient methods of safety practices with modern technology.

The algorithm intended for this project will be utilizing Proctor's PPO and discrete environment setup with controlled parameters [7] to test out training two agents. The reasoning behind starting with two agents is to conserve computational cost while generating training data, allowing the agents to "learn" their discrete environment. The location of each agent within the discrete environment is unimportant, as they will be communicating with one another utilizing signal to noise ratio [7] upon conclusion of their training episode.

Depending on the data gathered, the cumulative reward system established will determine which weights in the stochastic gradient move the agents closer to their target radiation source. Once desirable weights with higher reward yield are established, another episode is run and more training data is gathered. The training will be considered complete when the agents have consistently discovered the most efficient pathway to their target radiation source.

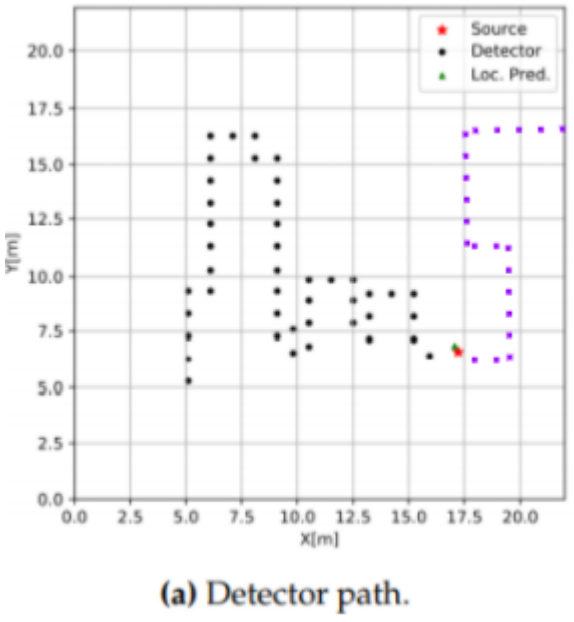

Fig 2: Modified 2D virtual environment as originally designed by Philippe Proctor [7], with an added purple dotted line signifying the second detector. Both agents will be accomplishing a shared objective in finding the radiation source (red star).

Once the agents complete their controlled 2D virtual environment as designed by Proctor [7], more agents can be conservatively added. When the algorithm has been tested to allow for multiples of agents, it can then be tested in a real world environment. This will be the final stage of testing the MARL algorithm before submitting it to its partnered research institution, the Defense Threat Reduction Agency (See **Acknowledgements**). They will be developing the complimentary hardware that the algorithm will be housed in.

This paper will examine multi-agent methodologies explored as well as approaches that seem to best match the foundations from which Proctor's single agent approach originates. Resources were limited due to the ongoing state of COVID-19, which impacted the project (which is described in **Methodology).** Due to time constraints, technological difficulties, and inability to successfully execute data during the life cycle of this research process, conclusive results for MARL are unavailable. See ***Discussion/Conclusions*** for further assessment and suggestions for next steps.

B. Multi-Agent Reinforcement Learning (MARL)

MARL combines a Multi-Agent System (MAS) with RL, allowing for completion of a shared objective with multiple available MAS strategies. [5]. With regards to Proctor's PPO approach in single agent radiation detection [7], an optimal strategy would be to utilize Differentiable Inter-Agent Learning (DIAL) [5]. This would allow for a faster assessment of radiation sources in urban environments without putting human bodies at risk.

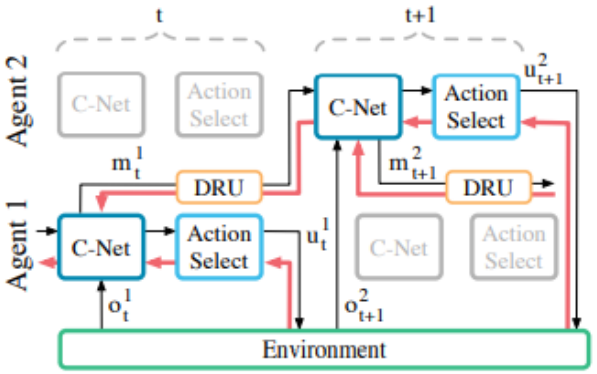

Fig. 3: A diagram of multi-agent reinforcement learning (MARL) showing how messaging bypasses the action selector through the Discretise/Regularise Unit (DRU), which passes as a continuous value to the next C-network (C-Net). [5]

C. Related Work

Several solutions have been investigated regarding the behavior of swarm algorithms and multi-agent based modeling. Kennedy and Eberhart [4] discuss the behavior of swarm algorithms, and how the agents would readjust their behavior based on the data collected for rewarded behavior in how they navigated their environment. The authors go on to discuss how they are able to trace travel patterns through binary numbers, and can track the shift by noting where the numbers change within the environment. It was also discovered that using discrete spaces helps to more easily discern patterns as repeated trials are made. This article has been referred to often as a foundational starting point for other RL papers, allowing for iterative designs to be made by later scientists studying agent based learning.

Liu and Abbaszadeh [3] implemented a deep Q-learning algorithm for radiation source search and compared it with a stochastic gradient search and uniform search method . They found that Q-learning is a more effective and efficient search algorithm. The literature goes into describing the foundations of RL, how the theory works, and then delves into proving said theory with trial runs. It was concluded that double Q-learning was preferred as a way to detect radiation in urban environments, and that for future iterations work would be done on drone-based radiation detection platforms

Schulman et al. [6] investigate policy gradient methods by comparing and testing different optimization procedures to understand what would best

support RL. The authors accomplished this through bounded stochastic gradient descent that prevented neural network weight updates from being too large, referred to as Proximal Policy Optimization (PPO). The authors compared PPO against five other policy optimization methods on a variety of RL environments.Through trial and error they were able to determine that PPO was the best out of all the other online policy gradient methods. It provided a reasonable combination of acceptable sample diversity in a straightforward manner within the defined parameters of the environment.

Foerster et al. [5] examine the complexities of communication between multiple agents through RL by examining two different approaches: Reinforced Inter-Agent Learning (RIAL) and Differentiable Inter-Agent Learning (DIAL). RIAL operates through Q-learning while DIAL functions by back-propogating error derivatives via noisy communication channels. [5]. They took an interesting approach that emphasizes centralized learning with a decentralized execution, meaning that the objective was shared but the agents could communicate their findings with one another. That way when training data was implemented in the next round of testing, the agents could retain previous experiences gathered independently to more efficiently navigate a discrete environment with set parameters collectively.

II. METHODOLOGY

A. Process

The goal for this research process was to build upon Proctor's pre-existing work with his single agent radiation detection algorithm and add more agents to locate radiation sources more quickly in his existing discrete localized environment. This was done using OpenAI [11], an open source research organization that studies machine learning through DRL. Their work specifically utilizes PPO, which was the chosen policy for Proctor's single agent radiation detection process.

After gaining familiarity with OpenAI's environment through their software tutorial SpinningUp, testing was done using a Windows 10 command terminal to access Portland State University's (PSU) Linux lab machines remotely. In person lab access was unavailable due to safety concerns regarding COVID-19. The coding language used was Python, as it is a more accessible programming tool for new researchers to access while testing the discrete environment. The Python IDE used locally on the Windows machine was PyCharm, as it responded better to troubleshooting instead of the more commonly used software, VisualStudio.

Python scripts were programmed by Proctor [7] to create a uniform search agent within a discrete environment for radiation detection. The objective was that the agent had to work for any starting detector position, so there needed to be an initial search direction that the agent could travel in. It wasn't necessary to explore the entire environment, and this setup was designed without obstructions for ease of testing.

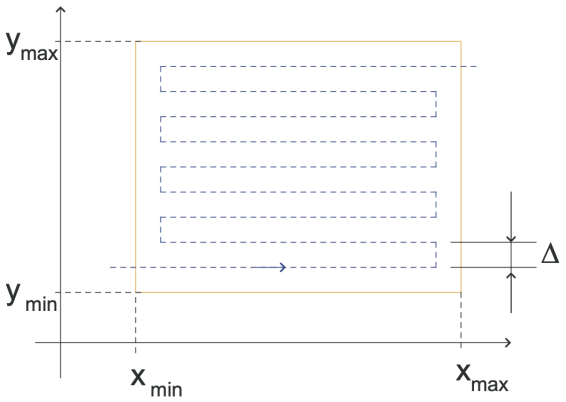

Fig. 4: Sample diagram of uniform search agent behavior if uniform_search.py was executed successfully

Due to complications with .py script availability and Anaconda package installation between the PSU lab machines and the local Windows 10 machine, the code was unable to be executed successfully. Each machine contained different elements of the proper scripts and packages separately, and attempted troubleshooting for PATH rectification and package installations were ultimately rejected by both machines.

B. Conceptual Approach

As has been stated throughout this paper, the objective of Multi-Agent Based Radiation Localization is to improve efficiency in radiation detection within urban environments to preserve human safety and health. This technological development is not intended to be a conclusive ending to development of radiation detection technology, but rather a significant milestone in the age of machine learning and DRL. Had the uniform search agent script been executed successfully with one agent,

the next logical step would be to add a second agent and observe their DRL outcomes.

A conceptual idea for how this project could be conducted successfully is to continue with the conservative approach of creating two agents to start. Then the two agents could be examined through the relationship between Forester et al's approach of DIAL [5] and Proctor's use of PPO [7]. Both research teams utilized elements of signal to noise ratio, which Proctor refined with Partially Observable Markov Decision Process (POMDP) [7] and Foerster et al. [5] exploited with presence of channel noise in the Discretise/Regularise Unit (DRU) for their learning model.

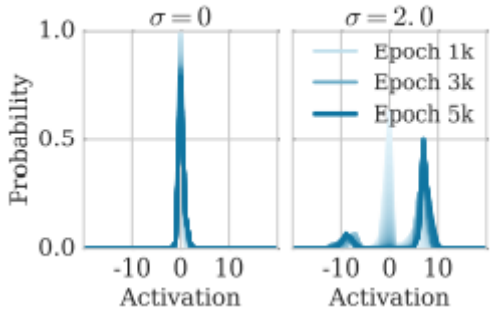

Fig 5.: DIAL experiment using the effect of channel noise for multi-agent reinforcement learning (MARL). The left graph shows centered activation of communication without noise, and the right graph shows how the presence of noise activates learning within the agents working for their shared objective. [5]

The presence of noise can be curtailed to the discrete parameters established in Proctor's original single agent environment, and both agents can explore their individual starting locations gathering data and adding learned weight values to their training. Upon gathering their information within their first episode and assessing the total reward, neural network weights can be updated and training begins again. This is done until both agents have determined the quickest route to the radiation source, storing learned information shared with one another to reduce unnecessary travel. Since Proctor's PPO [7] has proven to be more adaptable to new randomized environments, and Foerster et al's DIAL approach [5] broadly mimics human ways of communication for a shared objective, it would make sense to combine both strategies for successful Multi-Agent Based Radiation Localization

III. CONCLUSION

This paper investigated the efficacy of MARL using Philippe Proctor's master's thesis work [7] on single agent radiation localization using PPO and his single agent algorithm script. Due to technical difficulties throughout the life cycle of this REU project, developing a second agent within Proctor's pre-existing discrete environment proved unsuccessful. It was recommended that ongoing work continue to refine MARL, and to explore a conceptual idea regarding Differential Inter-Agent Localization (DIAL) [5] as a multi-agent tactic within PPO.

IV. DISCUSSION OF FUTURE WORK

Originally the vision for this project was to create more than two agents, and have a swarm algorithm designed to locate radiation sources. Upon concluding the REU life cycle of this project, it is recommended that ongoing development of MARL conservatively test two agents so as to create manageable checkpoints for algorithm development. The reason for this suggestion is to have better control of algorithm performance within the chosen coding script, and as errors are corrected more agents can be confidently added. Following successful algorithm testing, it is recommended that the algorithm be placed in autonomous platforms such as drones, robots, and/or radiation detection devices to test in more realistic environments to gauge real world response.

ACKNOWLEDGMENT

The author would like to thank their mentors Philippe Proctor and Dr. Christof Teuscher for their endless words of wisdom, support, and insight for such a complex project. The REU Site is supported by the National Science Foundation under grant no 1758006. The research was supported by the Defense Threat Reduction Agency (DTRA) under grant no HDTRA1-18-1-0009.

REFERENCES

[1] James Kennedy and Russell C. Eberhart. "A Discrete Binary Version of the Particle Swarm Algorithm." *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, October 12, 1997, 5. https://doi.org/10.1109/ICSMC.1997.637339.

[2] Hasselt, Hado van, Arthur Guez, and David Silver. "Deep Reinforcement Learning with Double Q-Learning." *ArXiv:1509.06461 [Cs]*, December 8, 2015. http://arxiv.org/abs/1509.06461.

[3] Liu, Zheng, and Shiva Abbaszadeh. "Double Q-Learning for Radiation Source Detection." *Sensors* 19, no. 4 (February 24, 2019): 960. https://doi.org/10.3390/s19040960.

[4] James Kennedy and Russell C. Eberhart. "A Discrete Binary Version of the Particle Swarm Algorithm." *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, October 12, 1997, 5. https://doi.org/10.1109/ICSMC.1997.637339.

[5] Foerster, Jakob N., Yannis M. Assael, Nando de Freitas, and Shimon Whiteson. "Learning to Communicate with Deep Multi-Agent Reinforcement Learning." *ArXiv:1605.06676 [Cs]*, May 24, 2016. http://arxiv.org/abs/1605.06676.

[6] Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. "Proximal Policy Optimization Algorithms." *ArXiv:1707.06347 [Cs]*, August 28, 2017. http://arxiv.org/abs/1707.06347.

[7] Philippe Proctor, Christof Teuscher, Adam Hecht, and Marek Osinski. "Proximal Policy Optimization for Radiation Source Search." *Sensors*, July 19, 2021, under review.

[8] Riley, Joshua, Radu Calinescu, Colin Paterson, Daniel Kudenko, and Alec Banks. "Reinforcement Learning with Quantitative Verification for Assured Multi-Agent Policies:" In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence*, 237–45. Online Streaming, --- Select a Country ---: SCITEPRESS - Science and Technology Publications, 2021. https://doi.org/10.5220/0010258102370245.

[9] Sutton, Richard S., and Andrew G. Barto. *Reinforcement Learning: An Introduction*. Second edition. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press, 2018.

[10] UnitedNations, ed. *Sources and Effects of Ionizing Radiation: United Nations Scientific Committee on the Effects of Atomic Radiation: UNSCEAR 2000 Report to the General Assembly, with Scientific Annexes*. New York: United Nations, 2000.

[11]Open AI. "OpenAI." Accessed August 18, 2021. https://openai.com/.