Exploration of Unknown Environments Using Deep Reinforcement Learning

Joseph McCalmon
Wake Forest University, Department of Computer Science
1834 Wake Forest Road
Winston-Salem, North Carolina 27109
mccajl18@wfu.edu

Abstract

My research presents a method for efficient exploration of an outdoor, unknown area, which aims to achieve precise coverage of regions of interest within that area. While this method for autonomous exploration was designed for autonomous controllers in unmanned aerial vehicles (UAVs), the concepts apply to any vehicle which uses autonomous navigation. We consider an environment with areas of interest of various sizes littered throughout, and a reinforcement learning agent which is tasked with discovering and mapping these areas in an efficient manner.

Introduction

Exploration of an unknown area is an important task in many applications of mobile robotics. Autonomous robots are employed in environmental mapping, detection of an area of interest, search and rescue missions, and other operations that involve navigating a previously unknown environment. Since these problems require constructing a map of a new environment while keeping track of the agent's location within it, they can be defined as Simultaneous Localization and Mapping (SLAM). Unmanned Aerial Vehicles (UAVs) are particularly effective in these tasks because recent advances in onboard computing power allow for increased maneuverability over ground robots. For these tasks, an accurate mathematical model of the environment is often unavailable, because the area of exploration is unknown in some capacity. Reinforcement learning (RL) in UAVs offers a solution to this problem because the RL agent does not need an explicit model of the environment to navigate within it; instead it learns the model by trial and error.

Many existing RL solutions with UAVs assume a target destination for the agent to reach, but many practical applications do not involve this form of explicit target. To our knowledge, no other work focuses on actually discovering and exploring areas of interest in a completely new area without predetermined targets. Building on the RL algorithms DDQN (Hausknecht and Stone 2015) , and A2C (Sutton and Barto 2018), we propose a model which can successfully navigate an infinitely large unknown area while seeking and following areas of interest with limited resources (i.e. time and battery life).

Research Methodology

For state of the art RL models, effective navigation of a large area is intractable because of the lack of a strong enough feedback signal. To combat this, we divide the map into many smaller regions. The first division is into nine equal regions, and then another division occurs which limits the agent to a 25x25 area, called the local map (Maciel-Pearson et al. 2019). The agent's vision size is a 5x5 area, with it in the center. By dividing the area the drone has to explore, we can also give it intermediary tasks to perform to make the problem solvable through RL algorithms. The first task is Target Selection. We use an optimization equation to pick the centroid of one of the nine regions as a temporary target for the agent to navigate towards, using information, gathered from previous time steps throughout the episode. This function considers the distance to that region, the percentage of that region that has been covered previously, and the percentage of the region that contained an area of interest, to as much precision as possible using the gathered information. With these metrics, the algorithm can select a promising, but relatively unknown region to explore.

Once a region is chosen, the agent transitions to another task called Target Search. The cell in the middle of each region is designated as the region target, and the goal of the target search task is to reach that target. Once again we run into the issue of a very large state space when trying to potentially navigate across multiple regions to the chosen target. To resolve this issue, we limit the agent's movement to the local map within its current assigned region, and designate a cell, closest in distance to the region target, as the local target. Hence, each search task consists of multiple smaller search tasks, where the agent reaches the local target and a new local map is formed around it. In addition, the agent is rewarded for covering any areas of interest along the way, so it learns to deviate slightly from the quickest path to the target when beneficial. We use the DDQN algorithm with an LSTM layer to implement this task

After reaching the region target, the agent transitions to Region Exploration. In this task, the agent is permitted to explore the entire region, and is rewarded for visiting unknown cells and covering areas of interest in minimal steps. The A2C algorithm is used to implement the Region Exploration.

Each testing episode for the RL agent consists of the cy-

Model	*30%*	*50%*	*70%*
Our Model	**3,874**	**7,959**	**18,674**
Curiosity	39,756	39,756	39,756
Zigzag	9,529	15,964	21,786
Random	14,640	33,817	40,245

Table 1: Average number of steps over 500 episodes to reach 30%, 50%, and 70% AoI coverage.

cle: select a target, navigate to that target, and explore the region surrounding that target. We let this repeat until the agent manages to cover 70% of the areas of interest, or $40,000$ steps, whichever comes first.

Results

Fig.1(b) shows an example path our RL agent took over the course of an episode to reach 70% area of interest coverage in the simulated map (Fig.1(a)). Fig.1(c) shows an agent performing the curiosity exploration as in (Burda et al. 2018). Fig.1(d) shows the path taken by a baseline of a hardcoded policy, which aims to simply cover as much ground as possible without overlapping, by sweeping from side to side. Fig.1(e) shows the path of the final baseline, an agent which uses the same proposed sub-tasks, but picks actions randomly instead of according to an RL policy.

Table 1 shows the average amount of steps each algorithm took to reach coverage thresholds of 30%, 50%, and 70%. Our proposed model manages to reach each of these thresholds faster than the three baselines. The curiosity exploration (Burda et al. 2018) mimics an RL agent without the architecture mentioned in the methodology section, which failed to reach any of the coverage thresholds in the allotted 40,000 steps. This shows our partitioning of the problem into subtasks is necessary for the RL algorithms to succeed. The sweeping policy as seen in Fig.1(d) performed closest to our model, but still took more steps because the learned policy in our model enables the UAV to efficiently search the entire map, rather than simply prioritizing covering every location. The random approach's inability to find AoI quickly shows that the RL algorithms are still necessary to take advantage of the problem decomposition. Ultimately, our model performs better than each baseline because it is trained to optimize information gain with each step, which is effective for missions in large, unknown environments.

Discussion and Future Work

As seen in the results, our agent explores large, unknown areas in less steps than the baselines. To extend our work to larger areas, we can simply partition into more, equally-sized regions. In the future, we will look extend the algorithm to a multi-agent environment, to further increase the area that can be explored.

We plan to work with the Peruvian National Park Service to deploy our algorithm onboard drones in the Amazon, to search for hotspots of illegal gold mining. Our agent supplies a method for efficiently exploring large regions which have not been imaged closely, and identify the locations of mining which damage the ecosystem. In addition, our algorithm can be applied to search and rescue (SaR) scenarios, as we have shown in our recent publication (McCalmon et al. 2020). The number of areas of interests can be altered to represent SaR targets, and the agent seeks them out quickly.

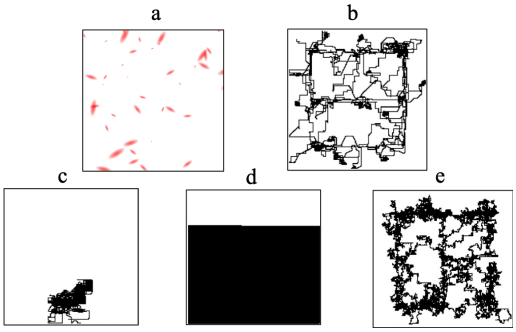

Figure 1: The original simulation map, with AoIs in red (a). Samples from the generated paths for the agent using the proposed approach (b), the curiosity exploration (Burda et al. 2018) (c), the hard coded sweeping policy (d), and the random policy (e).

References

Burda, Y.; Edwards, H.; Pathak, D.; Storkey, A. J.; Darrell, T.; and Efros, A. A. 2018. Large-Scale Study of Curiosity-Driven Learning. *CoRR* abs/1808.04355. URL http://arxiv.org/abs/1808.04355.

Hausknecht, M. J.; and Stone, P. 2015. Deep Recurrent Q-Learning for Partially Observable MDPs. *CoRR* abs/1507.06527. URL http://arxiv.org/abs/1507.06527.

Maciel-Pearson, B. G.; Marchegiani, L.; Akcay, S.; Abarghouei, A. A.; Garforth, J.; and Breckon, T. P. 2019. Online Deep Reinforcement Learning for Autonomous UAV Navigation and Exploration of Outdoor Environments. *CoRR* abs/1912.05684. URL http://arxiv.org/abs/1912.05684.

McCalmon, J.; Peake, A.; Zhang, Y.; Raiford, B.; and Alqahtani, S. 2020. Wilderness Search and Rescue Missions using Deep Reinforcement Learning. *SSRR* .

Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement Learning: An Introduction*. The MIT Press, second edition. URL http://incompleteideas.net/book/the-book-2nd.html.