Learning to Navigate in Unseen Cluttered Environment

Carnegie Mellon University
The Robotics Institute

Vidhi Jain†, Ganesh Iyer, Katia Sycara
{†vidhij, giyer, katia}@cs.cmu.edu

Problem

Learning end-to-end policy for navigation with a focus on intelligent exploration has been found to be a challenging task in embodied AI. While methods like soft-Q learning and ensembles of policies have demonstrated navigation behaviors in completely observed maps, we currently do not have ways of extending these policies to unexplored or partially explored environments. To this end, we propose a modular hierarchical formulation by decomposing the navigation task into two sub-problems: selecting the next best goal in the visible space, followed by efficiently navigating to this space in the partial map setting.

Intuition

When navigating in unknown environments, humans often invoke a decision criteria. This could be the next sub-goal, and the decision could either be based on some semantic context, or structural priors. Learning how to select the next sub-goal for higher level task such as coverage of the map is a challenging open question. We consider frontiers, which are the points at the boundary of known regions of a map, as points to facilitate exploration in a map. Therefore, we define such frontiers as sub-goals, and learn a policy that can select such frontiers on partial grid maps.

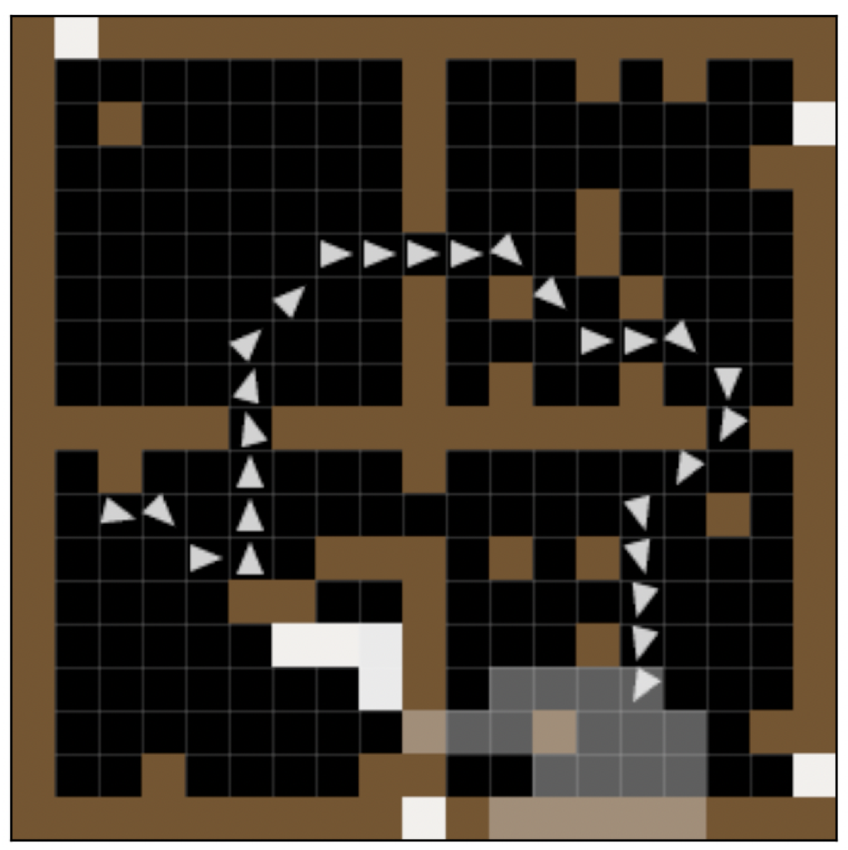

Fig. 1: FourRooms environment with clutter and human-like field of view

We demonstrate our approach in customized version of the classic FourRooms environment in gym-minigrid [3], with a realistic human-like field of view (shown in grey area in Fig 1), and variation in clutter density for evaluating the navigation policies.

Related Work

The concept of frontiers for exploration of unknown spaces was first introduced in [1] by Yamauchi. Related work in embodied AI literature has focused on the object and instance based navigation in 3D world environments. Our work is very close to [2] which demonstrated such a hierarchical policy for point- goal navigation tasks in 3D environments. Our approach differs in terms of being a study on performance of reinforcement learning algorithms to predict frontiers and contrast classical vs learning approaches for navigation with systematically varying the clutter density in the environment.

Approach

We propose a global policy network that learns the selection of the next best goal('frontier') in the observable space, followed by a local policy that learns low-level navigation conditioned on different sub-goal embeddings in known environments.

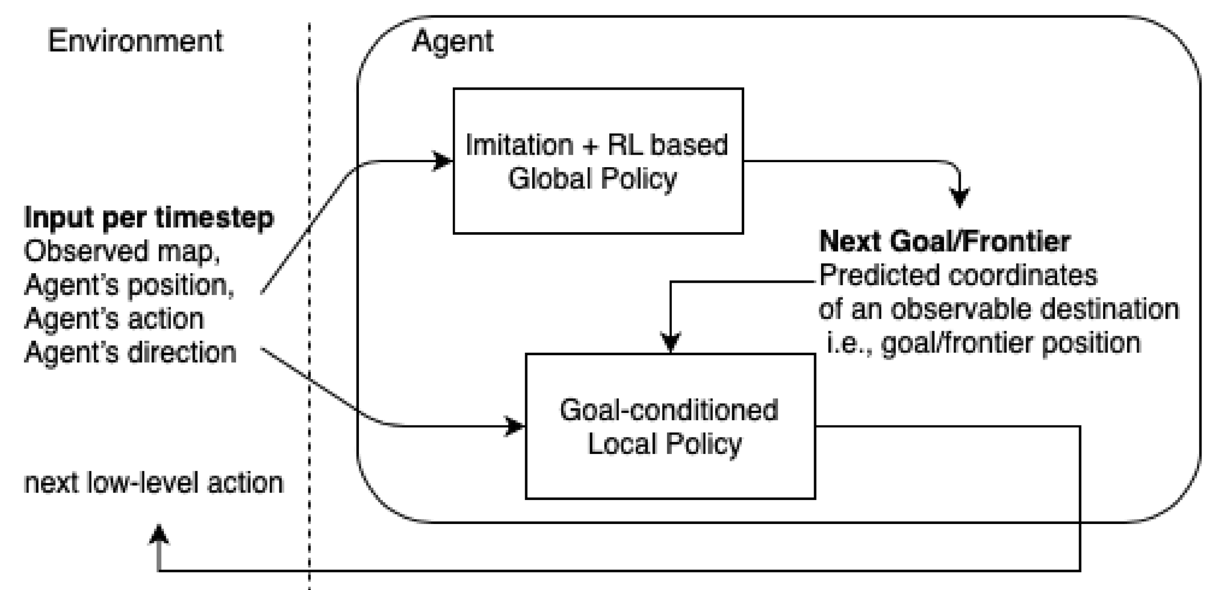

Fig. 2: Proposed Architecture

The global policy is trained on a continuous action space to predict the next sub-goal, and the local policy learns to take discrete actions like turn 15 degrees left/right and move forward. The input to both the policies is the belief over the observed map so far. Our local policy can be considered as a short term goal driven planner but trained as goal-conditioned RL. The combined use of planning and RL aligns with our global policy where we learn high-level goal selection policy by A^* planning to find the sequence of states in the path. We have trained both the policies with PPO, where the reward is based on unique area covered at each timestep for the global policy.

Clutter density is the probability of grid cell being impassable, in addition to the walls of the four rooms. The variations of clutter density in the Fourrooms environment are shown in Fig 3. Note that the increasing clutter density hinders the vision of the agent. The agent's trajectory is shown with white triangles. These snapshots are recorded for NFBE with A^* planner at 100 steps.

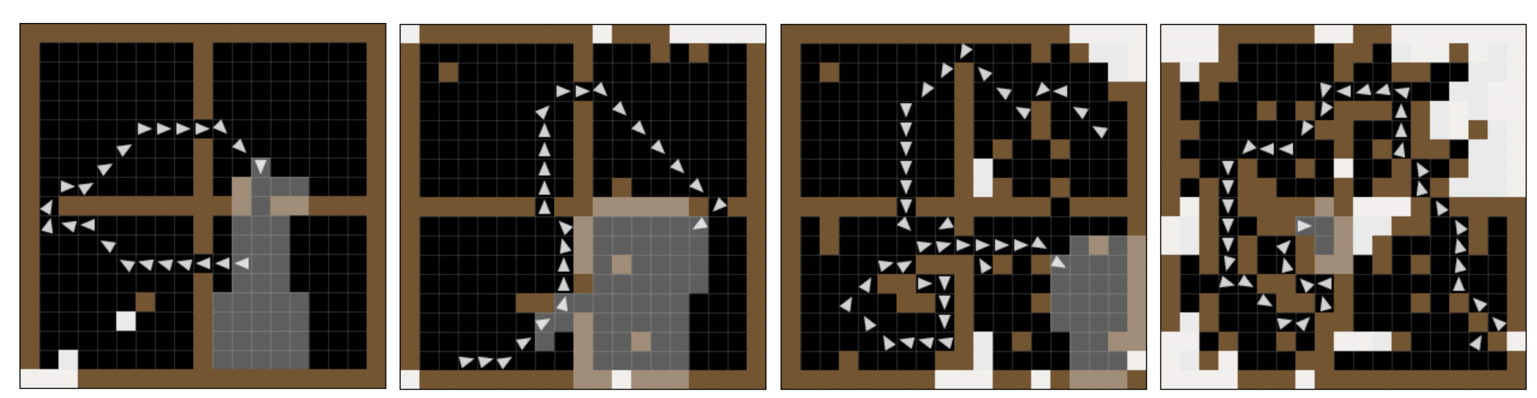

Fig. 3: clutter density (a) 0.01, (b) 0.03, (c) 0.1, (d) 0.3

References

[1] Brian Yamauchi. A frontier-based approach for autonomous exploration. In Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97. 'Towards New Computational Principles for Robotics and Automation', pages 146–151, 1997.
[2] Devendra Singh Chaplot, Dhiraj Gandhi, Saurabh Gupta, Abhinav Gupta, and Ruslan Salakhut- dinov. Learning to explore using active neural slam. In International Conference on Learning Representations, 2020.
[3] Maxime Chevalier-Boisvert, Lucas Willems, and Suman Pal, Minimalistic Gridworld Environment for OpenAI Gym, https://github.com/maximecb/gym-minigrid

Comparison

We compare our method with nearest frontier based exploration (NFBE) and end-to-end trained policies for coverage task. Although classic approaches perform well in the original FourRooms environments, our research question is to compare the learning and traditional navigation approaches when the clutter density in the rooms increases. The notion of clutter density is analogous to the real world case, where a lot of frontiers detected may occur in the small enclosed spaces, and requires learning of the structural priors for intelligent navigation.

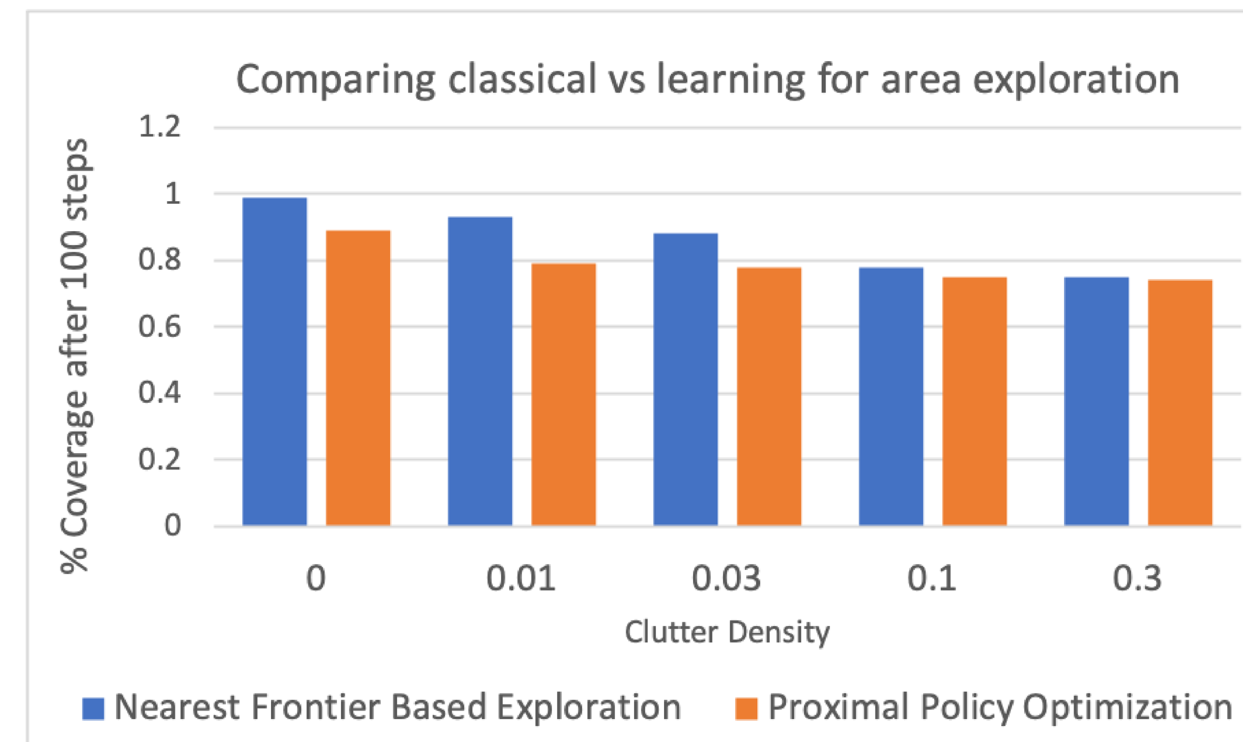

Fig. 4: Coverage $\%$ based on clutter density

We observe that the gap in coverage of nearest frontier based exploration (NFBE) and proximal policy optimization (PPO) for higher clutter is less than 1%. This suggests that learning based methods can perform at par with classical navigation techniques when the nearest frontier may not be the optimal one to maximize coverage in cluttered environments.

Remarks

Increasing the clutter density leads to impassable areas in the map and it becomes close to the traditional maze puzzle without much learnable structural information. So, we cannot arbitrarily increase the density in this comparison. The next step is to realistically organize clutter and incorporate movable clutter in the 3D environments, for example with indoor house simulators.

Learning based approaches are sensitive in terms of performance with the change in environments and show a large variance in the coverage percentage. End-to-end training of policy gradients algorithm is unstable due to non-stationary rewards based on coverage. Decomposing the low-level navigation to local policy while using the high-level structure information with the global policy allowed us to train the networks for comparison.

Acknowledgements

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001120C0036. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Defense Advanced Research Projects Agency (DARPA).