Generalizing Navigation Behaviors with Policy Sketches

Vihang Agarwal, David Fouhey

February 2019

1 Introduction

We plan to explore Navigation in realistic 3D environments and at large scales. Long-range 3D navigation is a complex cognitive task that relies on developing an internal representation of space, grounded by recognisable landmarks and robust visual processing. Our work will focus on representing a framework for decomposing navigation from building up internal representations of environments, planning or learning mobile operation on specialized tasks to a multi task reinforcement learning setting where learning a composition of behavioral subpolicies allows agent to solve tasks in these environments and generalize high-level behavior at large scales.

We would focus on methods for learning hierarchical policy representations. Practically it has been shown that discovery of these hierarchies requires detailed supervision in the form of explicitly specified high-level actions, subgoals, or behavioral primitives. This poses various questions on the nature of supervision necessary and a better model to achieve full benefits of hierarchies.

In this study we will try to use demonstrations to supervise discoveries of these hierarchies. Given policy demonstrations and a space of behavioral primitives to the agent, we hope that it be able to infer the underlying primitives as subpolicies in these demonstrations and use these subpolicies to generalize navigational ability over unseen navigational tasks and environments.

2 Related Work

3D Navigation and Generalization: There has been a prominent line of work on the task of navigation in real 3D scenes. Mirowski shows that an agent's navigation ability can be improved in mazes by introducing auxiliary tasks. However, these only evaluate the agent's generalization ability on pixel-level variations or small mazes. A more recent body of work on long range navigation relies on integration of general policies with locale-specific knowledge, and propose a dual pathway architecture that allows locale-specific features to be encapsulated, while still enabling transfer to multiple cities. Yet these works

rely on shaped rewards to embed navigating behaviors in the agent and train on static street view images rather than rich 3D scenes. (Gupta et al., 2017) [4] show that an agent can learn to navigate via mapping and planning using shortest path supervision in an end-to-end learning framework.

Instead, we will follow (Andreas et al., 2017) [2] where they explore a multitask reinforcement learning setting where the learner is presented with policy sketches (Policy sketches are short, ungrounded, symbolic representations of a task that describe its component parts. While symbols might be shared across tasks, the learner is told nothing about what these symbols mean, in terms of either observations or intermediate rewards) and (Pathak et al., 2018)[1] where an agent explores the environment without any expert supervision and distills this exploration data into goal-directed skills. These skills are then used to imitate the visual demonstration provided by some expert.

3 The Environment

We plan to use House 3D [5] as the virtual environment to test agents generalizibility to unseen navigation tasks and is sourced from SUNCG dataset. It has 45,622 human-designed 3D scenes ranging from single-room studios to multi-floor houses. On average, there are 8.9 rooms and 1.3 floors per scene. Each scene in SUNCG is fully annotated with 3D coordinates and its room and object types (e.g. bedroom, shoe cabinet, etc). At every time step an agent has access to the following signals: a) the visual RGB signal of its current first person view, b) semantic/instance segmentation masks for all the objects visible in its current view, and c) depth information. In House3D, an agent can live in any location within a 3D scene, as long as it does not collide with object instances (including walls) within a small range, i.e. robot's radius. Doors, gates and arches are considered passage ways, meaning that an agent can walk through those structures freely.

4 Task Formulation

Let the supervision demonstrations be available as a sequence of images $D : \{{x_1}^d, {x_2}^d, ..., {x_N}^d\}$. Also, let each demonstration be a sequence of subpolicies corresponding to a sequence of high-level symbolic primitives drawn from a fixed vocabulary B. We consider a multitask reinforcement learning problem as a partially observable markov decision process in a shared environment. This environment is defined by a tuple (S, A, P, γ). Let $S : \{x_1, a_1, x_2, a_2, ..., x_K\}$ be the sequence of observation and low-level actions generated by an agent as it explores its environment. We won't assume any prior on the number of low level actions available or how to use these actions, which must be inferred by the model. Each task $\tau \in T$ is specified by a reward R_τ which is a task specific reward function and the goal observation x_g to reach. Our aim is to learn a policy π that takes as input a pair of observations (x_i, x_g) and outputs a

sequence of subpolicies ($\pi_\tau : \pi_1, \pi_2, ..., \pi_K$) to reach its goal.

We will also use the paradigm of curriculum learning. Initially the model will be presented with task demonstrations associated with short sketches. Once average reward on all these tasks reaches a certain threshold, the length limit will be incremented.

5 Evaluation

The task for the model is to learn subpolicies for high level behaviors from demonstrations and use these subpolicies to generalize its navigational ability to large scale environments. Thus a natural question would be to ask: do these subpolicies actually allow the agent to learn these primitive behaviors? One evaluative criteria would be to see how well an agent can demonstrate a particular primitve behavior such as exiting a building, or moving north as far as possible etc. from these learned policies. It must also be able to exhibit compositionality, where behavior from a sequence of these subpolicies must be consistent with the expected actions that the agent takes.

For most navigational tasks, it is intuitive to think that reaching the goal is more important that how it is reached. Thus the agent must understand that a goal has been reached. The agent must show the ability to navigate without being lost. We will also use the geodesic distance [3] to measure proximity. This will also be a measure of the agent's ability to learn about shortcuts. It is important to think about shortcuts in embodied navigation. The agent must learn to take shortcuts in unseen navigation tasks and this behavior must be robust to changes in the environment.

References

[1] Guanghao Luo Pulkit Agrawal Dian Chen Yide Shentu Evan Shelhamer Jitendra Malik Alexei A. Efros Trevor Darrell Deepak Pathak, Parsa Mahmoudieh. Zero-shot visual imitation, 2018.

[2] Sergey Levine Jacob Andreas, Dan Klein. Modular multitask reinforcement learning with policy sketches, 2017.

[3] Devendra Singh Chaplot Alexey Dosovitskiy Saurabh Gupta Vladlen Koltun Jana Kosecka Jitendra Malik Roozbeh Mottaghi Manolis Savva Amir R. Zamir Peter Anderson, Angel Chang. On evaluation of embodied navigation agents, 2018.

[4] Sergey Levine Rahul Sukthankar Jitendra Malik Saurabh Gupta, James Davidson. Cognitive mapping and planning for visual navigation, 2017.

[5] Georgia Gkioxari Yuandong Tian Yi Wu, Yuxin Wu. Building generalizable agents with a realistic and rich 3d environment, 2018.