Robots that can remember like animals

Luke Bidulka

October 29, 2022

1 Introduction

Animals often know what kinds of environments they expect to encounter, and how they have successfully navigated them before. These past skills can be leveraged to face environments based on how similar a current environment looks to be to a past one. For example, if you start to drive in heavy snowfall, you will recognize this new environment for what it is and act in ways which worked well in similar past conditions. You would not immediately start re-learning how to drive in the snow. Similarly, artificial agents could re-use or start from policies learned for past environments and directly apply them in the present, allowing for quick adaptation to environments as they are encountered. Obviously, this is not a simple or straightforward computational task.

Inspired by the application of MAP-Elites to Robots that can adapt like animals and POET, I propose a method where an agent first recognizes the environment it finds itself in, then retrieves the most appropriate policy for this specific environment from an archive of pre-computed elite policies. The method consists of two parts:

1. Pre-Training: the environment is parameterized in select dimensions and optimization is run to find the elite agent policies (fixed agent morphology and network topology) in each discretized region (cell) of possible environments. Separately, a system-ID policy is trained to predict environment parameters from exploratory steps in the environment.

2. Deployment: The agent is deployed into a new environment and uses the learned system-ID policy for an exploration period, then predicts the environment parameters from this collected exploration data. The pre-computed elite for this corresponding cell of environment space is then retrieved and used moving forward.

2 Key Milestones & Sub-Objectives

1. Proof of concept on simple, already parameterized OpenAI gym Lunar Lander environment (has gravity, wind power, wind turbulence parameters):
   (a) Build an archive of elite PG agents for each environment parameter region cell using direct optimization and grid search
   (b) Train system-ID agent to predict the environment parameters from interactions in the environment
   (c) Evaluate:
      i. Performance of zero-shot elite application to current environment
      ii. Performance of fine-tuning elite to current environment
      iii. Performance against baseline PG trained on diverse environment parameters to determine overall gain via proposed method
      iv. Performance against control method of randomly selecting elites from the map instead of using system-ID agent

2. Perform 1 again, but this time with a POET-like method of elite map creation. Allowing the utilization of stepping stones to create elite agents for each environment parameter region cell. This would be expected to increase performance.