Empirical Evaluation of Hierarchical Reinforcement Learning

Rashmi Iyer1, 2, Zhiyu Zhang2, Ioannis Paschalidis2
The Harker School, 500 Saratoga Ave, San Jose, CA1;
Boston University, Center for Information and Systems Engineering, 15 St. Mary’s Street, Brookline, MA2

Introduction

Reinforcement learning (RL) is an area of machine learning where a computer learns how to complete a task in order to maximize a numerical reward. Unlike supervised learning, RL agents are not given a training set and instead learn through experience, discovering which actions help the agent yield the most reward by trial-and-error. The difficulty with traditional RL methods, however, is that the environment may not be efficiently explored if the reward is sparse or delayed. Additionally, the learned knowledge may not be easily adapted to similar situations. Recently, there has been a trend of applying hierarchical reinforcement learning (HRL) methods to sequential decision making processes. In HRL, a high-level controller is used to solve tasks by determining low-level subgoals that are then used to direct and train the low-level policy. Exploration is improved due to prolonged actions and intrinsic rewards, and learned knowledge can be shared. In a recent work (Nachum et al., 2019), a notion of sub-optimality of a subgoal representation is formalized and bounded. An algorithm is proposed to find near-optimal subgoal representations by optimizing the bound, and its performance is further tested on Mujoco Ant environments. The goal of this study is to verify the generalizability of this HRL algorithm on other deep RL benchmarks.

Methods

Using the Spinning Up framework and basic RL algorithm, a HRL program is developed and applied to the OpenAI Gym Cart Pole environment, a classical 2D control task. Various hyperparameters are then tuned, including the c-step (the number of steps that pass before the high level controller determines a new goal) and maximum episode length, and the algorithm is tested with varying random seed values.

Discussion/Conclusions

The results of this study support the idea that HRL methods can be applied to the Cart Pole environment to yield comparable and better results, especially at smaller c-steps and greater maximum episode lengths. These results suggest the potential advantage of HRL methods in future machine learning research. Future work would include testing the HRL methods on other environments such as Atari Games to confirm the generalizability of the algorithm.

References


Acknowledgements

I would like to thank Zhiyu Zhang for all of his guidance and help with this project, and I would like to thank Dr. Paschalidis and the BU RISE program for this opportunity.