SPEEDING UP POLICY OPTIMIZATION WITH VANISHING HYPOTHESIS AND VARIABLE MINI-BATCH SIZE

Anonymous authors
Paper under double-blind review

ABSTRACT

Reinforcement learning-based algorithms have been used extensively in recent years due to their flexible nature, good performance, and the increasing number of said algorithms. However, the largest drawback of these techniques remains unsolved, that is, it usually takes a long time for the agents to learn how to solve a given problem. In this work, we outline a novel method that can be used to drastically reduce the training time of current state-of-the-art algorithms like Proximal Policy Optimization (PPO). We evaluate the performance of this approach in a unique environment where we use reinforcement learning to help with a practical astronomical problem: where to place a fixed number of observatory stations in the Solar System to observe space objects (e.g. asteroids) as permanently as possible. That is, the reward in this scenario corresponds to the total coverage of the trajectories of these objects. We apply noisy evaluation for calculating the reward to speed up the training, which technique has already been efficiently applied in stochastic optimization. Namely, we allow the incorporation of some additional noise in the reward function in the form of a hypothesis term and a varying mini-batch size. However, in order to follow the theoretical guidelines, both of them are forced to vanish during training to let the noise converge to zero. Our experimental results show that using this approach we can reduce the training time remarkably, even by 75%.