Abstract

One of the main goals of artificial intelligence field is to solve complex tasks which have high-dimensional observation spaces. In recent year, remarkable progress has been made by combining deep learning and reinforcement learning to achieve high-level performance in the video and board games, 3D navigations and robotics control. In this thesis, deep reinforcement learning algorithms are studied to perform the robotic tasks with continuous action spaces.

Firstly, we use deep deterministic policy gradient (DDPG) combined with hindsight experience replay (HER) with very simple binary reward to achieve a multi-goal reinforcement learning task, learning the policy of reaching any given position for the redundant manipulator. Then we use DDPG with shaped reward to train redundant manipulator to fulfill the same task. By referring to the idea of hindsight experience replay, we propose future and random strategy to obtain some additional goals combined with shaped reward to generate some new transitions, which can help to promote the sample efficiency. After that, we use DDPG with prioritized experience replay to train a mobile robot learning to move along some specific trajectories. Two training strategies, random referenced state initialization and early termination are introduced to enable the mobile robot to learn effectively from the referenced trajectories.

Secondly, we study distributed deep reinforcement learning algorithms. We use asynchronous advantage actor-critic (A3C) and synchronous advantage actor-critic (A2C) algorithm, both of which have multiple workers to collect transitions and compute the gradients, to train the manipulator to complete a more complex task by adding some obstacles in the space. We also propose a new reward function to optimize the reaching path of end-effector. The performances of agents trained by different algorithms and reward functions are compared and their generalized abilities are also tested by changing the position and size of the obstacles. Next, we propose a distributed framework of DDPG, where the synchronous workers generate samples and compute gradients for the global network and the collecting workers only produce transitions to the shared replay memory with different policies. We use distributed DDPG to train the mobile robot and it could make the agent learn faster and achieve smaller tracking errors compared with single-worker DDPG.

Next, we focus on the proximal policy optimization algorithm (PPO) with generalized advantage estimation (GAE). We also propose a distributed framework for PPO by running multiple workers to collect transitions for global network at the same time. Then we use distributed PPO to train the manipulator learning to arrive at any given position and meanwhile avoid the obstacles. After that, we adopt three methods to train the mobile robot tracking the trajectories. The first method is to use distributed PPO with improved random referenced state initialization and early termination. In order to promote the training efficiency, the second method uses a two-stage training strategy consisting of supervised pre-training and fine-training by distributed PPO. For the third method, we introduce LSTM to represent the actor and critic, and use replay to store the cell state and hidden state of LSTM, which will be used for the initialization

of each episode to solve the problem of inaccurate LSTM initial states. The tracking performances of these three methods are also compared.

Finally, we utilize deep reinforcement learning to train the autonomous vehicle to learn driving behaviors by only taking the raw sensor data as input. We design a reward function which encourages the vehicle to drive along the road smoothly and also try to overtake other vehicles. We propose a two-stage training strategy: the vehicle is firstly trained on a simple road and then it continues to be trained on a more complicated road. The second training stage in a more complex environment can further improve the policy, leading to better driving behaviors. We use DDPG and improved DDPG, called TD3 to train the autonomous vehicle respectively and find TD3 could achieve better driving performance.