

Trust Region Policy Optimization - Implementation

Pietro Nardelli

Master's Degree in Artificial Intelligence and Robotics
Department of Computer, Control and Management Engineering
Sapienza University of Rome

January 2020

1 Introduction

Trust Region Policy Optimization is an algorithm that make several approximations to a theoretical iterative procedure for optimizing policies with guaranteed monotonic improvement. Despite its approximations that deviate from the theory, TRPO tends to give monotonic improvement, while little tuning of hyperparameters. This algorithm is effective for optimizing large nonlinear policies such as neural networks. The algorithm has been tested on two different openAI gym environments. Gym library is a collection of test problems that can be used to test reinforcement learning algorithms.

2 TRPO algorithm

3 Environments

3.1 MountainCarContinuous-v0

An underpowered car must climb a one-dimensional hill to reach a target. The action (engine force applied) is allowed to be a continuous value. The target is on top of a hill on the right-hand side of the car. If the car reaches it or goes beyond, the episode terminates.

On the left-hand side, there is another hill. Climbing this hill can be used to gain potential energy and accelerate towards the target. On top of this second hill, the car cannot go further than a position equal to -1, as if there was a wall. Hitting this limit does not generate a penalty.

The observations are CarPosition and CarVelocity and the only action permits to push the car on the left (negative values) or on the right (negative values).

Reward is 100 for reaching the target of the hill on the right hand side, minus the squared sum of actions from start to goal.

This reward function raises an **exploration challenge**, because if the agent does not reach the target soon enough, it will figure out that it is better not to move, and won't find the target anymore. To consider the problem solved, the reward should be over 90.

3.2 LunarLanderContinuous-v0

4 Results